

# PROGRAM CONTROLLER AND MEMORY ALLOCATION DEVICE AND METHOD

Publication number: JP2000099351 (A)

Also published as:

Publication date: 2000-04-07

JP3826626 (B2)

Inventor(s): HIRONO MITSUAKI; INUI KAZUYUKI; KONAKA YOSHIOHARU; KURIBAYASHI HIROSHI

Applicant(s): OMRON TATEISI ELECTRONICS CO

Classification:

- international: G06F12/00; G06F9/46; G06F9/48; G06F9/52; G06F12/00; G06F9/46; (IPC-1-7): G06F9/46; G06F12/00

- European:

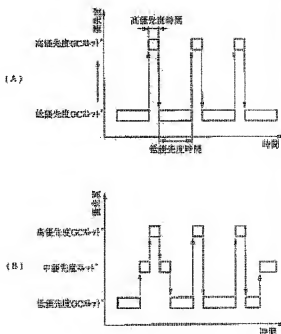
Application number: JP19990200345 19990714

Priority number(s): JP19990200345 19990714; JP19970321766 19971121

## Abstract of JP 2000099351 (A)

**PROBLEM TO BE SOLVED:** To attain an incremental CG even with a mark &amp; sweep method, to omit the compaction and also to improve the availability of a memory, to decrease the fragments in consideration of the lifetime of each object and to reduce the waste of power of a CPU.

**SOLUTION:** A means is prepared to incrementally perform a garbage collection thread which detects an object that is not referred to by any object in a heap area of a memory and releases the memory area of the detected object as a memory allocatable free area of another object together with a means which performs the scheduling to execute in time division each thread according to its priority and a means which alternately changes the priority of the garbage collection thread in both states where the priority is higher and lower than the priorities of other threads.



Data supplied from the [esp@cenet](mailto:esp@cenet) database — Worldwide

(19) 日本国特許庁 (J P)

## (12) 公開特許公報 (A)

(11) 特許出願公開番号

特開2000-99351

(P2000-99351A)

(43) 公開日 平成12年4月7日 (2000.4.7)

(51) Int.Cl. <sup>7</sup>	識別記号	F I	メモコード (参考)
G 0 6 F 9/46	3 4 0	G 0 6 F 9/46	3 4 0 B
			3 4 0 F
12/00	5 9 1	12/00	5 9 1

審査請求 未請求 請求項の数34 O L (全30頁)

(21) 出願番号 特願平11-200345  
 (52) 分割の表示 特願平10-260427の分割  
 (22) 出願日 平成10年9月14日 (1998.9.14)  
 (31) 優先権主張番号 特願平9-321766  
 (32) 優先日 平成9年11月21日 (1997.11.21)  
 (33) 優先権主張国 日本 (J P)

(71) 出願人 000002945  
 オムロン株式会社  
 京都府京都市右京区花園土堂町10番地  
 (72) 発明者 廣野 光明  
 京都府京都市右京区花園土堂町10番地 オムロン株式会社内  
 (72) 発明者 乾 和志  
 京都府京都市右京区花園土堂町10番地 オムロン株式会社内  
 (74) 代理人 100094548  
 弁理士 小森 久夫

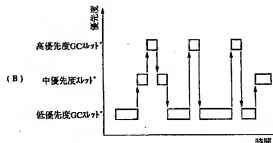
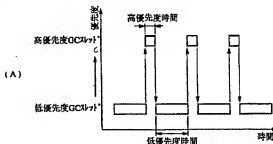
最終頁に続く

## (54) 【発明の名称】 プログラム制御装置とメモリ割当装置および方法

## (57) 【要約】

【課題】 マーク&スイープ法でありながらインクリメンタルなGCを実現する。コンパクションを不要とし、且つメモリの使用効率を向上させる。オブジェクトごとの寿命を考慮してフラグメントを低減し、メモリの使用効率を向上させ、CPUパワーの無駄を削減する。

【解決手段】 メモリのヒープ領域内の、どのオブジェクトからも参照されないオブジェクトを検出し、そのメモリ領域を他のオブジェクトのメモリ割り当て可能なフリー領域として解放するガベージコレクションスレッドをインクリメンタルに行う手段と、スレッドの優先度に応じて各スレッドを時分割に実行するためのスケジューリングを行う手段と、前記ガベージコレクションスレッドの優先度をガベージコレクションスレッド以外のスレッドの優先度より高い状態と低い状態とに交互に変更する手段とを設ける。



## 【特許請求の範囲】

【請求項1】 或るスレッドからの、指定メモリ領域に対するデータの書き込み有無検出の開始を依頼するアプリケーションプログラムインタフェースの呼び出しにตอบสนองして、その書き込み有無を示すフラグを書き込み無し

の状態に設定する手段と、  
前記指定メモリ領域に対するデータの書き込みがあったとき、前記フラグを書き込み有りの状態に設定する手段と、前記スレッドからの、指定メモリ領域に対するデータの書き込み有無検出の終了を依頼するアプリケーション  
10 プログラムインタフェースの呼び出しにตอบสนองして、前記フラグの状態に応じた値を前記スレッドに返す手段とを備えたことを特徴とするプログラム制御装置。

【請求項2】 メモリのヒープ領域内の、どのオブジェクトからも参照されないオブジェクトを検出し、当該オブジェクトのメモリ領域を他のオブジェクトのメモリ割り当て可能なフリー領域として解放するガベージコレクションスレッドをインクリメンタルに行う手段と、スレッドの優先度に応じて各スレッドを時分割に実行するためのスケジューリングを行う手段と、前記ガベージコレクションスレッドの優先度をガベージコレクションスレッド以外のスレッドの優先度より高い状態と低い状態とに交互に変更する手段とを設けたことを特徴とするプログラム制御装置。

【請求項3】 前記ガベージコレクションスレッドの優先度の高い状態の時間をアプリケーションプログラムインタフェースの呼び出しによって設定する手段を設けたことを特徴とする請求項2に記載のプログラム制御装置。

【請求項4】 前記ガベージコレクションスレッドの優先度の高い状態と低い状態による周期をアプリケーションプログラムインタフェースの呼び出しによって設定する手段を設けたことを特徴とする請求項2に記載のプログラム制御装置。

【請求項5】 前記フリー領域の容量の時間変化を検出する手段と、前記フリー領域の容量が低下傾向のとき、前記ガベージコレクションスレッドの優先度の高い状態の時間を長くする手段を設けたことを特徴とする請求項2または3に記載のプログラム制御装置。

【請求項6】 イベントの発生に応じてリアルタイムスレッドを実行させ、該リアルタイムスレッドの中断時または終了時に非リアルタイムスレッドを実行させる手段を備えたプログラム制御装置において、前記非リアルタイムスレッドの1つは、メモリのヒープ領域内の、どのオブジェクトからも参照されないオブジェクトを検出し、当該オブジェクトのメモリ領域を他のオブジェクトのメモリ割り当て可能なフリー領域として解放するガベージコレクションをインクリメンタルに行うガベージコレクションスレッドであり、当該ガベージ  
40 コレクションスレッド以外の非リアルタイムスレッドの

実行によりヒープ領域内の前記フリー領域が所定の値まで減少したとき、前記ガベージコレクションスレッドを実行する手段を設けたことを特徴とするプログラム制御装置。

【請求項7】 メモリのヒープ領域内の、どのオブジェクトからも参照されないオブジェクトを検出し、当該オブジェクトのメモリ領域を他のオブジェクトのメモリ割り当て可能なフリー領域として解放する、手順の異なる複数通りのガベージコレクションスレッドを選択的に実行する手段と、前記フリー領域または前記オブジェクトの使用領域の量に基づいて、前記複数通りの手順のうち1つの手順のガベージコレクションスレッドを実行する手段を設けたことを特徴とするプログラム制御装置。

【請求項8】 メモリのヒープ領域内に割り当てられるオブジェクトのサイズの分布を検知する手段と、前記分布の中心より大きな固定サイズの整数倍を前記フリー領域に対する新たなオブジェクトの割り当てサイズに決定する手段とを設けたことを特徴とするメモリの割り当て装置。

【請求項9】 前記オブジェクトのサイズの分布を検知する手段は、システムの使用開始時にシステムに組み込まれ、前記オブジェクトのサイズの分布を検知した後にはシステムから切り離されるプログラムモジュールによるものである請求項8に記載のメモリの割り当て装置。

【請求項10】 前記固定サイズを決定する手段は、システムの使用開始時にシステムに組み込まれ、前記固定サイズが決定された後にシステムから切り離されるプログラムモジュールによるものである請求項8に記載のメモリの割り当て装置。

【請求項11】 前記ヒープ領域内に予め複数のサイズに分割するヒープ領域分割手段と、オブジェクト生成時に当該オブジェクトのサイズより大きく且つ最も小さなサイズの領域を割り当てする手段を設けたことを特徴とする請求項8に記載のメモリの割り当て装置。

【請求項12】 前記ヒープ領域分割手段は、前記複数のサイズの分割数と各サイズを指数とするアプリケーションプログラムインタフェースの呼び出しによって、前記ヒープ領域を分割するものである請求項11に記載のメモリの割り当て装置。

【請求項13】 メモリのヒープ領域内にオブジェクトを生成する手段を備えたプログラム制御装置において、固定サイズの整数倍を前記ヒープ領域内に対するオブジェクトの割り当てサイズに決定する手段と、アプリケーションプログラムインタフェースの呼び出しによって前記固定サイズを設定する手段とを設けたことを特徴とするメモリの割り当て装置。

【請求項14】 メモリのヒープ領域内にオブジェクトを生成する手段を備えたプログラム制御装置において、固定サイズの整数倍を前記ヒープ領域内に対するオブジェクトの割り当てサイズに決定する手段と、

前記ヒープ領域内に割り当てられるオブジェクトのサイズの分布をアプリケーションプログラムインタフェースの呼び出しによって設定する手段と、該アプリケーションプログラムインタフェースの呼び出しにตอบสนองして、前記分布の中心より大きな値を前記固定サイズとして設定する手段とを設けたことを特徴とするメモリの割り当て装置。

【請求項 15】 テンプレートとしてのクラスと該クラスにより生成されるオブジェクトをメモリのヒープ領域内に割り当てるメモリの割り当て装置において、前記クラスからオブジェクトが生成された時刻に相当するデータを記憶し、前記オブジェクトを削除する際に、当該オブジェクトの寿命を検出し、該寿命のデータをクラスに設け、当該クラスからオブジェクトを生成する際に、前記寿命のデータに基づいて、前記ヒープ領域に対するオブジェクトの生成領域を分ける手段を設けたことを特徴とするメモリの割り当て装置。

【請求項 16】 メモリのヒープ領域内の、他のオブジェクトから参照されるオブジェクトを検出して当該参照の有無状態を記憶する参照情報記憶手段と、当該参照情報記憶手段の記憶内容を基に、どのオブジェクトからも参照されないオブジェクトのメモリ領域を他のオブジェクトのメモリ割り当て可能な領域として解放するオブジェクト消去手段を設けたメモリ割り当て装置において、前記参照情報記憶手段は、オブジェクトの参照関係を表すツリー構造の第 1 のデータと、オブジェクトの参照関係の変更された部分のオブジェクトを示す第 2 のデータとを記憶し、第 1 のデータを探索して参照されているオブジェクトを検出し、第 2 のデータを読み出すとともに該データを基に第 1 のデータを探索して、参照されてい

るオブジェクトを検出して、参照オブジェクト検出手段を設けたことを特徴とするメモリの割り当て装置。

【請求項 17】 前記参照オブジェクト検出手段は、前記オブジェクトの参照関係の変更時の参照先が前記第 1 のデータの探索により初めて検出されたオブジェクトであるときにのみ、参照先の当該オブジェクトを前記第 2 のデータとして記憶するものである請求項 16 に記載のメモリの割り当て装置。

【請求項 18】 或るスレッドからの、指定メモリ領域に対するデータの書き込み有無検出の開始を依頼するアプリケーションプログラムインタフェースの呼び出しにตอบสนองして、その書き込み有無を示すフラグを書き込み無しに状態に設定するステップと、前記指定メモリ領域に対するデータの書き込みがあったとき、前記フラグを書き込み有りの状態に設定するステップと、前記スレッドからの、指定メモリ領域に対するデータの書き込み有無検出の終了を依頼するアプリケーションプログラムインタフェースの呼び出しにตอบสนองして、前記フラグの状態に応じた値を前記スレッドに返すステップとから成るプログラム制御方法。

【請求項 19】 メモリのヒープ領域内の、どのオブジェクトからも参照されないオブジェクトを検出し、当該オブジェクトのメモリ領域を他のオブジェクトのメモリ割り当て可能なフリー領域として解放するガベージコレクションスレッドをインクリメンタルに行うステップと、スレッドの優先度に応じて各スレッドを時分割に実行するためのスケジューリングを行うステップと、前記ガベージコレクションスレッドの優先度をガベージコレクションスレッドに交互の優先度の優先度より高い状態と低い状態とに交互に変更するステップとから成るプログラム制御方法。

【請求項 20】 イベントの発生に応じてリアルタイムスレッドを実行させ、該リアルタイムスレッドの中断時または終了時に非リアルタイムスレッドを実行させるステップを備えたプログラム制御方法において、前記非リアルタイムスレッドの 1 つは、メモリのヒープ領域内の、どのオブジェクトからも参照されないオブジェクトを検出し、当該オブジェクトのメモリ領域を他のオブジェクトのメモリ割り当て可能なフリー領域として解放するガベージコレクションをインクリメンタルに行うガベージコレクションスレッドであり、当該ガベージコレクションスレッド以外の非リアルタイムスレッドの実行によりヒープ領域内の前記フリー領域が使用される値まで減少したとき、前記ガベージコレクションスレッドを実行するステップを備えたことを特徴とするプログラム制御方法。

【請求項 21】 メモリのヒープ領域内の、どのオブジェクトからも参照されないオブジェクトを検出し、当該オブジェクトのメモリ領域を他のオブジェクトのメモリ割り当て可能なフリー領域として解放するとともに、前記フリー領域または前記オブジェクトの使用領域の量に基づいて、手順の異なる複数通りのガベージコレクションスレッドを選択的に実行することを特徴とするプログラム制御方法。

【請求項 22】 メモリのヒープ領域内に割り当てられるオブジェクトのサイズの分布を検出し、その分布中心より大きな固定サイズの整数倍を前記ヒープ領域に対する新たなオブジェクトの割り当てサイズとすることを特徴とするメモリの割り当て方法。

【請求項 23】 テンプレートとしてのクラスと該クラスにより生成されるオブジェクトをメモリのヒープ領域内に割り当てるメモリの割り当て方法において、前記クラスからオブジェクトが生成された時刻に相当するデータを記憶し、前記オブジェクトを削除する際に、当該オブジェクトの寿命を検出し、該寿命のデータをクラスに設け、当該クラスからオブジェクトを生成する際に、前記寿命のデータに基づいて、前記ヒープ領域に対するオブジェクトの生成領域を分けることを特徴とするメモリの割り当て方法。

【請求項 24】 メモリのヒープ領域内の、他のオブ

5

エクトから参照されるオブジェクトを検出して当該参照の有無状態を記憶し、当該記憶内容に基づき、どのオブジェクトからも参照されないオブジェクトのメモリ領域を他のオブジェクトのメモリ割り当て可能な領域として解放するメモリ割り当て方法において、オブジェクトの生成時にオブジェクトの参照関係を表すツリー構造の第1のデータを記憶し、参照関係の変更時に、オブジェクトの参照関係の変更された部分のオブジェクトを示す第2のデータを記憶し、第1のデータを探索して参照されているオブジェクトを検出し、第2のデータを読み出すとともに該データを基に第1のデータを探索して、参照されているオブジェクトを検出することを特徴とするメモリの割り当て方法。

【請求項25】 或るスレッドからの、指定メモリ領域に対するデータの書き込み有無検出の開始を依頼するアプリケーションプログラムインタフェースの呼び出しにตอบสนองして、その書き込み有無を示すフラグを書き込み無し状態に設定する処理プログラムと、

前記指定メモリ領域に対するデータの書き込みがあったとき、前記フラグを書き込み有りの状態に設定するステップと、前記スレッドからの、指定メモリ領域に対するデータの書き込み有無検出の終了を依頼するアプリケーションプログラムインタフェースの呼び出しにตอบสนองして、前記フラグの状態に応じた値を前記スレッドに返す処理プログラムとを記録して成るプログラム記録媒体。

【請求項26】 メモリのヒープ領域内の、どのオブジェクトからも参照されないオブジェクトを検出し、当該オブジェクトのメモリ領域を他のオブジェクトのメモリ割り当て可能なフリー領域として解放するガベージコレクションスレッドをインクリメンタルに行う処理プログラムと、スレッドの優先度に応じて各スレッドを時分割に実行してスケジューリングを行う処理プログラムと、前記ガベージコレクションスレッドの優先度をガベージコレクションスレッド以外のスレッドの優先度より高い状態と低い状態とに交互に変更する処理プログラムとを記録して成るプログラム記録媒体。

【請求項27】 イベントの発生に応じてリアルタイムスレッドを実行し、該リアルタイムスレッドの中断時または終了時に非リアルタイムスレッドを実行するプログラム制御装置に用いられるプログラム記録媒体であって、

前記非リアルタイムスレッドの1つを、メモリのヒープ領域内の、どのオブジェクトからも参照されないオブジェクトを検出し、当該オブジェクトのメモリ領域を他のオブジェクトのメモリ割り当て可能なフリー領域として解放するガベージコレクションをインクリメンタルに行うガベージコレクションスレッドとし、当該ガベージコレクションスレッド以外の非リアルタイムスレッドの実行によりヒープ領域内の前記フリー領域が所定の値まで減少したとき、前記ガベージコレクションスレッドを実

6

行する処理プログラムを記録したことを特徴とするプログラム記録媒体。

【請求項28】 メモリのヒープ領域内の、どのオブジェクトからも参照されないオブジェクトを検出し、当該オブジェクトのメモリ領域を他のオブジェクトのメモリ割り当て可能なフリー領域として解放する、手順の異なる複数通りのガベージコレクションスレッドを選択的に実行する処理プログラムと、前記フリー領域または前記オブジェクトの使用領域の量に基づいて、前記複数通りの手順のうち1つの手順のガベージコレクションスレッドを実行する処理プログラムとを記録して成るプログラム記録媒体。

【請求項29】 プログラムの実行時にメモリのヒープ領域内にオブジェクトを生成するコンピュータに用いられるプログラム記録媒体であって、

メモリのヒープ領域内に割り当てられるオブジェクトのサイズの分布を検出し、その分布の中心より大きな固定サイズの整数倍を前記ヒープ領域に対する新たなオブジェクトの割り当てサイズとする処理プログラムを記録して成るプログラム記録媒体。

【請求項30】 前記ヒープ領域内を予め複数のサイズに分割し、オブジェクト生成時に当該オブジェクトのサイズより大きく且つ最も小さなサイズの領域を割り当てて成る処理プログラムを記録して成る請求項29に記載のプログラム記録媒体。

【請求項31】 メモリのヒープ領域内にオブジェクトを生成する手段を備えたプログラム制御装置に用いられるプログラム記録媒体であって、アプリケーションプログラムインタフェースの呼び出しによって前記固定サイズを設定する処理プログラムと、前記固定サイズの整数倍を前記ヒープ領域内に対するオブジェクトの割り当てサイズに決定する処理プログラムを記録して成るプログラム記録媒体。

【請求項32】 メモリのヒープ領域内にオブジェクトを生成する手段を備えたプログラム制御装置に用いられるプログラム記録媒体であって、前記ヒープ領域内に割り当てられるオブジェクトのサイズの分布をアプリケーションプログラムインタフェースの呼び出しによって設定し、該アプリケーションプログラムインタフェースの呼び出しにตอบสนองして、前記分布の中心より大きな値を前記固定サイズとして設定する処理プログラムと、前記固定サイズの整数倍を前記ヒープ領域内に対するオブジェクトの割り当てサイズに決定する処理プログラムを記録して成るプログラム記録媒体。

【請求項33】 プログラムの実行時にテンプレートとしてのクラスによりメモリのヒープ領域にオブジェクトを生成するコンピュータに用いられるプログラム記録媒体であって、

前記クラスからオブジェクトが生成された時刻に相当するデータを記憶し、前記オブジェクトを削除する際に、

50

当該オブジェクトの寿命を検出し、該寿命のデータをクラスに設け、当該クラスからオブジェクトを生成する際に、前記寿命のデータに基づいて、前記ヒープ領域に対するオブジェクトの生成領域を分ける処理プログラムを記録して成るプログラム記録媒体。

【請求項34】 メモリのヒープ領域内の、他のオブジェクトから参照されるオブジェクトを検出して当該参照の有無状態を記憶し、当該記憶内容に基づきオブジェクトを消去するコンピュータに用いられるプログラム記録媒体であって、

オブジェクトの生成時にオブジェクトの参照関係を表すツリー構造の第1のデータを記憶し、参照関係の変更時に、オブジェクトの参照関係の変更された部分のオブジェクトを示す第2のデータを記憶し、第1のデータを探索して参照されているオブジェクトを検出し、第2のデータを読み出すとともに該データを基に第1のデータを探索して、参照されているオブジェクトを検出する処理プログラムを記録して成るプログラム記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】 この発明は、コンピュータのプログラム制御装置、メモリ割当装置およびそれらの方法に関する。

【0002】

【従来の技術】 先ず、本願発明の明細書において用いる各種用語の定義を以下に示す。尚、出典記号は次のとおりである。

【0003】 JIS: JIS工業用語大辞典第4版(日本規格協会刊)

参1: 情報技術用語辞典(オーム社刊)

参2: 情報処理用語大辞典(オーム社刊)

参3: 先端ソフトウェア用語辞典(オーム社刊)

参A: SUPER ASCII Glossary Help on Internet

(1) プロセス(process): 処理や過程を意味する一般用語(参2)

(2) スレッド(thread): プロセスまたはタスクと呼ばれる1つの実行環境の中で並列実行可能な処理を複数に分割した場合に、プロセスのスケジューリング対象となる実行単位または制御フローのこと。(参3)

(3) コンテキスト(context): (オブジェクト指向システムに関連して) メソッドを実行するために必要な情報を蓄えるオブジェクトをいう。コンテキストは、呼び出し先のコンテキスト、プログラムの入ったメソッドオブジェクト、プログラムカウンタ、スタックポインタという情報、引数や一時変数を取るところ、および評価スタックから成立する。このような実行環境をオブジェクトとしてとるが、プロセスなどをサポートする高級言語の特徴でヒープ言語と呼ばれる。他方PASCALやALGOL60では実行環境はスタックにとられFORTRANでは固定領域にとられる。(参1)

(4) タスク(task): 多重プログラミング又は多重プロセッシングの環境において、計算機によって実行されるべき仕事の要素として制御プログラムによって取り扱われる命令の1つ以上の列(JIS)

(5) ガベージ(garbage): どこからも参照されていないが、生成されてしまったオブジェクト。ガベージはガベージコレクションで回収する。(参1)

(6) ガベージコレクション(Garbage Collection): (オブジェクト指向システムに関連して) メモリ管理の中心となるプログラムを言う。主記憶を使いきるとメインの計算を止め、ガベージコレクションを動かして、もはや使われていないオブジェクトを全部集める方法である。この方法でガベージコレクションが働くこと、計算が止まってしまうため入出力が全く反応しなくなり、実時間の応答が必要な用途には使えない。(参1)

(7) インタプリタ(interpreter): 解釈実行を行う翻訳プログラム(JIS)

(8) リアルタイム(real time): 計算機外部の他の処理との関係をもちながら、かつ外部の処理によって定められる時間要件に従って、計算機の行うデータの処理に関する用語。実時間。(JIS)

(9) オブジェクト(object): プロシージャ(定義した手続き)とデータの特性を結合させるエンティティ(実体)であり、これにより計算が行われ、局所的状態が蓄えられる。(参1)

(10) クラス(class): 同一の手続き群とデータ構造を持つオブジェクトの集合。(参3)

(11) ヒープ領域(heap area): プログラム実行時に必要に応じて使用されるような作業領域。(参2)

(12) スケジューリング(scheduling): ディスパッチされるべきジョブ又はタスクを選択すること。(JIS)

(13) イベント(event): ハードウェア/ソフトウェアが、自分自身の状態の変化を他のハードウェア/ソフトウェアに通知すること。一般にこの際の通知では、イベントの種類やハードウェア/ソフトウェアの状態を表す各種パラメータをメッセージとしてまとめ、相手に送信する。そしてイベントの通知を受けた側では、メッセージのパラメータなどから適切な処理を行う。(参A)

(14) イベントフラグ(event flag): タスクが1つまたは複数の事象の発生を待ち合わせるための機能とその事象を通知する機能とからなるタスク間同期通信機構。(参3)

(15) セマフォ(semaphore): 複数のプロセスやタスクを並列に処理するシステムで、各プロセス間、タスク間の同期やメッセージ制御、割込処理を行うための仕組み。(参3)

(16) 仮想マシン(仮想機械)(virtual machine): 複数の特定のプラットフォーム(OSやハードウェア)に組み込まれた、特定のプラットフォームに依存しないアプリケーションプログラムを実行する環境。同じ仮想マシン

ンさえ提供されていれば、プラットフォームが変わっても同じアプリケーションプログラムが実行できる。

【0004】(17) Java VM (Java Virtual Machine) : オペレーティングシステムに組み込まれた、Javaアプリケーションプログラムを実行するための環境。一般的なプログラムは、ソースコードをコンパイルして、それぞれのオペレーティングシステムに最適化した実行コードを生成する、というスタイルを採る。Javaで書かれたプログラムも同様の手順で作成するが、コンパイル後のコードは、特定のオペレーティングシステムに依存しない中間コードの形をとる。これをロードし、各オペレーティングシステムに合わせたコードに変換しながら実行するのがJava仮想マシンの役目であり、同じ仮想マシンさえ提供されていれば、プラットフォームが変わっても同じコードが実行できるようにしている。

【0005】(18) フリー領域(free area) : ヒープ領域上の使用可能な領域、未使用領域。

【0006】(19) ノーマルスレッド(normal thread) : リアルタイム性が要求されない処理を行うスレッド。

【0007】(20) マークテーブル(mark table) : オブジェクトがどこからも参照されないか否かを調べるために存在するオブジェクトに1対1に対応する表。或るオブジェクトに参照があることが確かめられた場合、そのオブジェクトに対応する表の欄にマークを付ける。全ての参照関係を調べたときマークの無いオブジェクトは不要であるので取り除くことができる。

【0008】(21) オブジェクトの寿命(life time of the object) : オブジェクトが生成され、消去されるまでの時間。

【0009】(22) ライトバリア(write barrier) : オブジェクトへの参照関係の変更をチェックし、書き替えが起こった場合に何らの処理をする。本件の場合には、書き替えが起きたとき、書き替える参照が指しているオブジェクトに対応するマークテーブルの欄にマークを付ける。

【0010】(23) スweep(sweeping) : ヒープ領域上の不要なオブジェクトを取り除く処理。

【0011】(24) オブジェクト生成(create the object) : 新しいオブジェクトを生成すること。具体的には、ヒープ領域の一部をオブジェクトに割り当て、内容を初期化すること。

【0012】(25) オブジェクト消去(delete the object) : 不要なオブジェクトを取り除くこと。具体的には、ヒープ領域に確保してある領域を解放すること。

【0013】(26) 参照(reference) : 或るオブジェクトAが別の特定のオブジェクトBにアクセスするためにオブジェクトBを特定する情報。具体的には、オブジェクトBを指すポインタまたはインデックス。

【0014】(27) 参照変更(change the reference /

reconnect the reference) : 参照を現在のオブジェクトBから別のオブジェクトCに変更すること。

【0015】さて、従来のシングルスレッドのコンピュータシステムにおいて、オペレーティングシステム上で複数のスレッドを並行処理する場合、共有メモリを用いて排他制御するために、また複数のタスク間で同期をとるために、従来よりセマフォやイベントフラグなどを用いて排他制御が行われていた。

【0016】また、例えばプログラムを少量のメモリ環境下で動作させることなどを目的として、仮想記憶を行わずに単一のアドレス空間のメモリをプログラムの実行時に動的に割り当て、いわゆる動的記憶管理が従来より行われている。このような動的記憶管理によれば、プログラムにより明示的にメモリ領域の解放を行わなければ、プログラムの実行過程において使用されなくなったメモリ領域が発生する。その結果、プログラムが使用できるフリー領域が次第に不足する。こうした問題を回避するために、使用されなくなったメモリ領域(ガベージ)を抽出し、これらを集めて(コレクトして)再び使用可能なフリー領域とする、ガベージコレクションと呼ばれる処理を自動的に行うようにしている。

【0017】ここで従来のガベージコレクションの処理手順をフローチャートとして図5に示す。ガベージコレクションのアルゴリズムにはマーク&スイープ法、複写法、参照カウント法などの各種方法が考えられているが、ここではマーク&スイープ法について例示する。図5.4に示すように、まずガベージコレクションの途中でガベージコレクションスレッド以外の他のスレッドが実行されないように割り込みを禁止し、シングルスレッドモードにする(s201)。続いてメモリ上のガベージコレクションの対象となる領域(以下「ヒープ領域」という。)に割り当てられているオブジェクトにそれぞれ対応するマークの記憶領域(以下「マークテーブル」という。)をクリアする(s202)。続いてヒープ領域内に割り当てられているオブジェクトの参照関係を示す情報を基に何れかのオブジェクトから参照されているオブジェクトを検出し、それらに対応するマークテーブル上の位置にマークを付ける処理を行う(s203)。この処理により、何れのオブジェクトからも参照されていないオブジェクトはもはや使われなくなったオブジェクトであり、それに相当するマークテーブルにはマークが付けられないことになる。従ってそのマークしていないオブジェクトを新たなオブジェクトの割当可能な領域、すなわちフリー領域として抽出する(s204)。例えばこのフリー領域をリスト構造のデータとして生成する。その後、割り込み禁止を解除し、マルチスレッドのモードに戻す(s205)。

【0018】従来はこのようなガベージコレクションが、メモリのフリー領域が所定量まで減少した時点で自動的に起動されるようにしていた。

【0019】また、ガベージコレクションを行ったと

11

き、任意の大きさ（メモリサイズ）のオブジェクトが任意に解放されるのでフラグメントが発生する。そこで、連続したサイズの大きな領域がとれるように、オブジェクトの割当領域を例えば先頭から順次詰めるメモリコンパクション（以下単に「コンパクション」という。）を実行していた。

【0020】

【発明が解決しようとする課題】 上述した排他制御の機能セマフォやイベントフラグなどのコンピュータ資源を用いることで実現している従来のシステムにおいては、その資源が他のスレッドなどで使用されている場合には、他のスレッドはその資源が解放されるのを待たなければならない。このような資源待ちの時間が生じると、リアルタイム性の要求されるシステムにおいては大きな障害となる。すなわち、資源待ち状態にあるスレッドは、その資源が解放されるまで、処理できずリアルタイムな応答が不可能となる。

【0021】 例えば上記従来のガベージコレクション

（以下「GC」という。）の方法によれば、メモリ空間が広いほど、ガベージとしてコレクトしてもよい領域が見つかるのに時間がかかり、例えば64～128MBのヒープ領域で数秒間を要し、且つGCはフリー領域がある程度減少した時点で不定期に行われるので、リアルタイム性の要求されるシステムには用いることができなかった。

【0022】 リアルタイム性の要求されるシステムでは、あるタスクを実行中に何らかのイベント（割り込み）が発生すれば、そのイベントに応じた他のスレッドを処理することになるが、そのスレッドの切替に要する時間は、最悪値として例えば数十μsec以下であることが要求される。ところが、上述したようにGCが何時起動されるか予測できず、一旦起動されれば、CPUは数秒間GCに専念することになるため、その間リアルタイム処理は不可能となる。

【0023】 上述の問題はGCの場合に限らず、長時間の資源待ち状態が生じるシステムに共通の問題である。

【0024】 この発明の目的はセマフォやイベントフラグなどのコンピュータ資源をロックのメカニズムとして使用せずに、処理の排他性を保証することにより、上述の問題を解消することにある。

【0025】 GCの他の方法として、従来の複写法をインクリメンタルに行うようにして、リアルタイム性を確保する方法が情報処理Vol. 35 No. 11 pp1008～pp1010に示されている。この方法によれば、GCの途中で中断を許すことができるので、時分的に他のスレッドと並行処理することが一応はできる。しかし、この複写法をインクリメンタルに行うようにする方法では、複写中に他のスレッドがメモリを使用する訳にはいかないで、メモリを使用しない極一部のスレッドしか並行処理できない。また、複写法では、複写元と複写先のメモリ領域を

12

確保しておかなければならないので、メモリの使用効率が低く、少量のメモリ環境下で動作させるシステムには向かない。

【0026】 また、マーク&スイープ法でマルチスレッドに対応するために参照関係が変更される度にマークを付与する方法は、「On-the-fly GC」という名前で、GC専用のCPUに処理を割り当てるマルチCPU化のための方法が情報処理Vol. 35 No. 11 pp1006～pp1008に示されている。しかし、この方法では、オブジェクトが常に表示されて、且つ参照関係が変更されている場合に、既にマークしているツリーを何回も辿り直して新しいノードを見つけてマークしなければならない、マーク作業がいつまでも終わらない場合が生じたり、非常に長い時間がかかる、という問題があった。また、この方法ではスイープ中にシステムをロックする必要があった。

【0027】 そこで、この発明の他の目的はマーク&スイープ法でありながらGCスレッドを他のスレッドと実質的に並行処理可能とし、GCの任意の時点で中断しても短時間に確実にGCを完了できるようにしたインクリメンタルGCを可能とすることにある。

【0028】 また、従来のマーク&スイープ法等のGCによるフラグメントを解消するためにコンパクションを行う際、多大なCPUパワーを必要としていた。また、このコンパクションを行わなければメモリの使用効率が低下するという問題があった。

【0029】 この発明の他の目的はコンパクションを不要とし、且つメモリの使用効率を向上させることにある。

【0030】 さらに、従来のオブジェクト指向システムにおいては、永続的に存在する寿命の長いオブジェクトと短時間で消滅する寿命の短いオブジェクトがヒープ領域内に混在していたため、GCを行うと、短寿命のオブジェクトから消去された後に確実にフラグメントが生じてメモリ使用効率が急激に低下する。また、GCを行う場合に、永続的に存在するオブジェクトに対してはもガベージであるか否かの判定を毎回行わなければならないので、このことがCPUパワーの無駄となっていた。これに対して、一定期間存在していたオブジェクトを永久に存在するものと見なし、そのオブジェクトを存在調査から外す、いわゆる世代別ガベージコレクションという方法も従来から考えられているが、このような方法でも不要なオブジェクトが存続したり、それぞれのオブジェクトについて一定期間の存続調査をしなければならないという欠点があった。

【0031】 この発明の他の目的はオブジェクトごとの寿命を考慮してフラグメントを低減し、メモリの使用効率を向上させ、CPUパワーの無駄を削減することにある。

【0032】 ここで、先行技術調査を行った際に発見した文献と本願発明との関係について示しておく。



# [0033] (1) Incremental Garbage Collection of Concurrent Objects for Real-Time Application

この論文はBaker が書いたという1978年のリアルタイムGCに対して、全体の処理時間から必要なGCの処理時間の割合を求めるものである。本願発明に係る課題を解決するものではない。

# [0034] (2) Distributed Garbage Collection for the Parallel Inference Machine: PIE64

GCを行う領域を細かく分けることにより、個々の処理時間を短くしてリアルタイム性を向上する方法である。本願発明のように全ての領域を対象にするものとは基本的に異なる。また、スケジューリングについては述べられていない。

# [0035] (3) Garbage Collection in Distributed Environment

Baker の考えを発展させ、ネットワークの分散環境に適用したもので、ネットワーク固有の問題を解決しようとするものであり、本願発明とは異なる。この分散環境の考え方に、本願発明のスケジューラを組み合わせることであり、分散環境でより高度なリアルタイムCGを実現することも可能となる。

# [0036] (4) 特開平1-220039号

システムコール発行時に、そのシステムコールにより起動されるタスクの優先度を制御するもの。優先度制御という点で共通点があるだけ。

# [0037] (5) 特開平3-231333号

オブジェクトのサイズを検出し、そのサイズに応じてワークエリアをメモリに確保することが一応示されている。

# [0038] (6) 電子情報通信学会誌Vol.80 No.6 pp58 6-592

マルチメディアオペレーティングシステムのコンセプトが示されている。

# [0039] (7) 日本ソフトウェア科学会第12回D7-4

スタック上のオブジェクトの一番上にオブジェクトの大きさを書いておくことが示されている。すなわちオブジェクトのサイズを記憶しておく、という点でのみ関連がある。

# [0040] (8) 11TH Real-Time System Symposium "Incremental Garbage Collection of Concurrent Object for Real-Time Applications"

オブジェクト間の参照ツリーに類似の考え方が示されている。

# [0041] (9) 情報処理学会第38回全国大会5U-7

オブジェクトの寿命に関連してメモリ割当を行うことが示されている。

# [0042] (10) Lecture Notes in Computer Science 259 "Garbage Collection in a Distributed Environment"

アプリケーションプログラムインタフェースの呼出に応じてGCを制御することと、オブジェクト参照に関する技術思想が示されている。

# [0043]

【課題を解決するための手段】請求項1, 18, 25に係る発明は、或るスレッドからの、指定メモリ領域に対するデータの書き込み有無検出の開始を依頼するアプリケーションプログラムインタフェースの呼び出しに応答して、その書き込み有無を示すフラグを書き込み無し状態に設定し、前記指定メモリ領域に対するデータの書き込みがあったとき、前記フラグを書き込み有りの状態に設定する。そして前記スレッドからの、指定メモリ領域に対するデータの書き込み有無検出の終了を依頼するアプリケーションプログラムインタフェースの呼び出しに応答して、前記フラグの状態に応じた値を前記スレッドに返す。

【0044】これにより、コンピュータ資源をロックのメカニズムとして使用せずに、処理の排他性が保証される。すなわち指定メモリ領域に対するデータの書き込み有無検出の開始を依頼するAPIの呼び出しから指定メモリ領域に対するデータの書き込み有無検出の終了を依頼するAPIの呼び出しまでの間で行われたスレッドAの処理の途中で、他のスレッドによる指定メモリ領域に対する書き込みがあったか否かが、そのスレッドAで分かる。書き込みが行われていなければ、指定メモリ領域の排他性が保たれている。もし書き込みが行われていれば、スレッドAのその間の処理を無効にして、例えばその処理を再度実行するなどの方法によって高い応答性を保ちながら、排他制御を行うことができるようになる。

【0045】請求項2, 19, 26に係る発明は、メモリのヒープ領域内の、どのオブジェクトからも参照されないオブジェクトを検出し、当該オブジェクトのメモリ領域を他のオブジェクトのメモリ割当て可能なフリー領域として解放するGCスレッドをインクリメンタルに行うスレッドを、他のスレッドの優先度より高い状態と低い状態とに交互に変更する。

【0046】このようにGCスレッドの優先度を交互に切り替えることによって、GCスレッドの優先度が低い状態では、他のスレッドによるアプリケーションが優先され、動作するアプリケーションが無い場合にはGCスレッドが自動的に動かされ、メモリのフリー領域が自動的に拡大される。GCスレッドの優先度が高い状態では、他のスレッドは実行されないが、優先度の低い状態と、他のスレッドにより、GCが継続して行われないことによってフリー領域が慢性的に不足状態となることが防止され、常に高いパフォーマンスを維持することができ。

【0047】請求項3に係る発明は、前記ガベージコレクションスレッドの優先度の高い状態の時間をアプリケーションプログラムインタフェースの呼び出しによって

設定する。これにより、APIを呼び出す側で、たとえばシステムの応答性を重視する場合には優先度の高い状態の時間を短くし、システム全体のスループットを重視する場合には優先度の高い状態の時間を長くする、といった設定が可能となり、システムのニーズに応じて性能を変えることができるようになる。

【0048】請求項4に係る発明は、前記ガベージコレクションスレッドの優先度の高い状態と低い状態による周期をアプリケーションプログラムインタフェースの呼び出しによって設定する。これにより、たとえば連続的なパルスを処理するような、連続してCPU資源が必要なアプリケーションプログラムの場合には、周期を短くし、通常のアプリケーションプログラムのような断続的にCPU資源を利用するアプリケーションプログラムでは周期を長くすることによって、コンテキストスイッチなどのオーバーヘッドを小さくすることができる。

【0049】請求項5に係る発明は、前記フリー領域の容量の時間変化を検出し、前記フリー領域の容量が低下傾向のとき、前記ガベージコレクションスレッドの優先度の高い状態の時間を長くする。これにより、動的にガベージコレクションの優先度の高い状態の時間を調整することになり、どのようなアプリケーションプログラムでもフリー領域が足りなくなるといった状況を回避することができる。且つフリー領域が十分確保できるシステムやアプリケーションプログラムでは、ガベージコレクションを最低限に抑えることができる。

【0050】請求項6、20、27に係る発明は、イベントの発生に応じてリアルタイムスレッドを実行させ、該リアルタイムスレッドの中断時または終了時に非リアルタイムスレッドを実行させるようにし、GCスレッド以外の非リアルタイムスレッドの実行によりフリー領域が所定の値まで減少したとき、非リアルタイムスレッドの1つであるGCを実行させる。

【0051】一般にリアルタイム性の要求されるシステムではリアルタイム性の必要なスレッドとそうでないスレッドが存在していて、一般にリアルタイム性が要求されるプログラムは生成するオブジェクトの量が少なく、その量を予測して設計することができる。一方、リアルタイム性の要求されないプログラムは、生成されるオブジェクトの量の予測が困難である。そこで、リアルタイム性の要求されるプログラムが生成すると予想されるオブジェクトの量を定義し、リアルタイム性の要求されないスレッドがオブジェクトを生成して、メモリのフリー領域の量が例えば上記定義したオブジェクトの量の近傍にまで減少した時に、その時点でスケジューリングを行い、GCスレッドを実行させる。このことによって、フリー領域が直ちに確保され、リアルタイム性の要求されるスレッドの実行可能な環境を常に維持することができる。

【0052】請求項7、21、28に係る発明は、メモ

リのヒープ領域内の、どのオブジェクトからも参照されないオブジェクトを検出し、当該オブジェクトのメモリ領域を他のオブジェクトのメモリ割り当て可能なフリー領域として解放するとともに、前記フリー領域または前記オブジェクトの使用領域の量に基づいて、手順の異なる複数通りのガベージコレクションスレッドを選択的に実行するようにする。

【0053】一般に、GCを行う場合、そのアルゴリズムに応じて、CPUに要求されるパワー、メモリの使用量、GCに要する時間などがそれぞれ異なる。そのため、フリー領域の量に応じて、適切なGCの手順は異なる。上述したように、フリー領域または使用領域の量に応じてGCの手順を切り替えることによって、常に効率的なGCが可能となる。

【0054】請求項8、22、29に係る発明は、メモリのヒープ領域内に割り当てられるオブジェクトのサイズの分布を検出し、その分布中心より大きな固定サイズの整数倍を前記ヒープ領域に対する新たなオブジェクトの割り当てサイズとする。

【0055】一般に、オブジェクト指向のシステムにおいては、プログラムの実行時に生成されるオブジェクトの大きさ（メモリ上のサイズ）の発生頻度分布は正規分布を示す。一方、新たなオブジェクトを生成する際、メモリ上のフリー領域から割り当て可能な領域を抽出するが、上述のようにオブジェクトのサイズの分布中心より、大きなサイズを新たなオブジェクトの割り当てサイズとしておくことによって、その割り当てられたオブジェクトが消去された後、新たなオブジェクトが割り当てられるとき、上記割り当てサイズより小さなオブジェクトであれば再利用できる。上記割り当てサイズは正規分布の分布中心より大きなサイズであるため、多くのオブジェクトが先に使用されていったメモリ領域を再利用することになる。また、このことにより、コンパクションを行わなくてもメモリの使用効率が向上することになる。また、コンパクションのためのCPUパワーが不要となり、小規模のCPUで応答性の高いシステムを構成することができる。

【0056】請求項9に係る発明は、前記オブジェクトのサイズの分布を検出する手段は、システムの使用開始時にシステムに組み込まれ、前記オブジェクトのサイズの分布を検出した後にシステムから切り離されるプログラムモジュールによるものとする。また、請求項10に係る発明は、前記オブジェクトの割り当てサイズを決定する手段は、システムの使用開始時にシステムに組み込まれ、前記オブジェクトの割り当てサイズが決定された後にシステムから切り離されるプログラムモジュールによるものとする。

【0057】これにより、一度オブジェクトのサイズの分布が検知された後は、また、オブジェクトの割り当てサイズが決定された後は、メモリ領域とCPUパワーの

無駄がなくなる。

【0058】請求項11、30に係る発明は、メモリのヒープ領域内にオブジェクトを生成する手段を備えたプログラム制御装置において、前記ヒープ領域内を予め複数のサイズに分割し、オブジェクト生成時に当該オブジェクトのサイズより大きく且つ最も小さなサイズの領域を割り当てる。

【0059】請求項12に係る発明は、前記複数のサイズの分割数と各サイズを指数とするアプリケーションプログラムインタフェースの呼び出しによって、前記ヒープ領域を分割する。

【0060】請求項13、31に係る発明は、メモリのヒープ領域内にオブジェクトを生成する手段を備えたプログラム制御装置において、固定サイズの整数倍を前記ヒープ領域内に対するオブジェクトの割り当てサイズに決定し、アプリケーションプログラムインタフェースの呼び出しによって前記固定サイズを設定する。

【0061】また、請求項14、32に係る発明は、メモリのヒープ領域内にオブジェクトを生成する手段を備えたプログラム制御装置において、固定サイズの整数倍を前記ヒープ領域内に対するオブジェクトの割り当てサイズに決定するものとし、前記ヒープ領域内に割り当てられるオブジェクトのサイズの分布をアプリケーションプログラムインタフェースの呼び出しによって設定し、該アプリケーションプログラムインタフェースの呼び出しに応答して、前記分布の中心より大きな値を前記固定サイズとして設定する。

【0062】これにより、別の装置、システム、およびアルゴリズムで、メモリのヒープ領域内に割り当てられるオブジェクトのサイズの分布を測定して、決定したオブジェクトの割り当てサイズの基になる固定サイズを登録することができ、CPUパワーとメモリの使用効率を向上させることができる。

【0063】請求項15、23、33に係る発明は、テンプレートとしてのクラスと該クラスにより生成されるオブジェクトをメモリのヒープ領域内に割り当てる際、前記クラスからオブジェクトが生成された時刻に相当するデータを記憶し、前記オブジェクトを削除する際に、当該オブジェクトの寿命を検出し、該寿命のデータをクラスに格納し、当該クラスからオブジェクトを生成する際に、前記寿命のデータに基づいて、前記ヒープ領域に対するオブジェクトの生成領域を分ける。

【0064】一般にオブジェクト指向のシステムにおいては、クラスをテンプレートとしてオブジェクトが生成されるため、同一のクラスから生成されたオブジェクトは略同一の寿命を持つ。そこで、或るクラスからオブジェクトが生成された時刻に相当するデータを記憶し、そのオブジェクトが削除された際に、当該オブジェクトの寿命を検出して、これをそのオブジェクトを生成したクラスの寿命データとして記憶し、そのクラスからオブ

ジェクトを生成する際に、寿命データに基づいて、異なるヒープ領域にオブジェクトを生成することによって、長寿命のオブジェクトと短寿命のオブジェクトが異なった領域に生成される。これにより、長寿命領域でのフラグメントが大幅に低下し、メモリの使用効率が向上する。また、GCを行う場合に、短寿命のオブジェクトが生成される領域を重点的に処理することによって、GCに消費されるCPUパワーを低減することができるようになる。

【0065】請求項16、24、34に係る発明は、メモリのヒープ領域内の、他のオブジェクトから参照されるオブジェクトを検出して当該参照の有無状態を記憶し、当該記憶内容に基づき、どのオブジェクトからも参照されないオブジェクトのメモリ領域を他のオブジェクトのメモリ割り当て可能な領域として解放するが、オブジェクトの生成時にオブジェクトの参照関係を表すツリー構造の第1のデータ（この第1のデータは、たとえばオブジェクトの参照関係を示す、オブジェクト内に設けられる1つの変数である。）を記憶し、参照関係の変更時に、オブジェクトの参照関係の変更された部分のオブジェクトを示す第2のデータを記憶し、第1のデータを探索して参照されているオブジェクトを検出し、第2のデータを読み出すとともに該データを基に第1のデータを探索して、参照されているオブジェクトを検出する。

【0066】このように構成することにより、スワイプやマークテーブルのクリアを行うときに、他のスレッドを止めなくても、スワイプ中またはマークテーブルのクリア中に他のスレッドにより新たなオブジェクトが生成されたり、オブジェクト間の参照関係が変更されても、マークテーブルを基にスワイプを行う際に、上記の新たに生成されたオブジェクトを消去してしまうことがない。そのため、インクリメンタルにGCを行うことができ、また、自由に中断ができるので、リアルタイム性が向上する。また、常にバックグラウンドでGCスレッドを動作させておくことが可能となり、メモリの使用効率を常に高く維持できる。しかも、ツリー探索によるマーク付与に要する時間が短縮化され、割込によって、いつまでもマーク付与が完了しない、といった不都合が防止でき、GCを確実に実行できる。

【0067】請求項17に係る発明は、前記参照オブジェクト検出手段は、前記オブジェクトの参照関係の変更時の参照先が前記第1のデータの探索により初めて検出されたオブジェクトであるときにのみ、参照先の当該オブジェクトを前記第2のデータとして記憶する。

【0068】これにより、オブジェクトの参照関係の変更時の参照先が既にマークされている場合に、第2のデータを重ねて記憶することがなくなり、第1のデータの探索およびマークに要する時間をその分短縮化できる。

【0069】

【発明の実施形態】この発明の実施形態であるプログ

ラム制御装置およびメモリ割り当て装置の構成を図1～図5を参照して順次説明する。

【0070】第1の実施形態 図1は装置のハードウェアの構成を示すブロック図である。装置は基本的にCPU1とオブジェクト群を生成するヒープ領域およびマテーブル等を記憶するメモリ2、および外部との入出力を行うI/O3とから構成する。また、外部から必要なプログラムをメモリにロードする場合は、CD-R OM読取インタフェース4を用い、プログラムが予め書き込まれたCD-ROM5を読み取るようにする。このCD-ROMが本願発明に係るプログラム記録媒体に相当する。

【0071】図2はソフトウェアの構成を示すブロック図である。同図において、カーネル部分はCPUやメモリを資源として管理し、時分割によるマルチスレッドの機能を実現する。VM (仮想マシン) 部分はプログラムとカーネルとのインタフェースを行うソフトウェアであり、ここではアプリケーションプログラムから見てVM以下の階層全体が例えばJava仮想マシンとして作用させる。(JavaはSun Microsystems社の商標) この場合、カーネルとVM部分がJavaOSを構成する。このVMにはプログラムがバイトコード等の中間コードで与えられるとき、これを解釈するインタプリタとその解釈に応じて呼び出されるプログラムモジュール等を含む。同図におけるプログラムは中間コードによる各種スレッドであり、上記インタプリタを介して、内部のプログラムモジュールを実行する。

【0072】図3はメモリのヒープ領域に生成されるオブジェクトの参照関係およびスタックとの関係を示す図である。ヒープ領域に対してオブジェクトを生成した際、或るオブジェクトから他のオブジェクトへの参照関係は同図に示すようにルートノードから延びるツリー構造を成す。例えばグローバル変数を宣言すれば、その変数に対応するオブジェクトが生成される。またスレッド毎に引き数領域、戻り番地、ローカル変数、作業領域等の情報を記憶するスタックが生成され、例えば図中矢印で示すようなスタック上のローカル変数からツリー上のグローバル変数への参照関係なども記憶される。これらのスタックはヒープ領域外の所定領域に格納される。

【0073】図4は図2に示したソフトウェアの構成をブロック図として詳細に示したものである。同図において、GCモジュール10はGCを行うための各種処理のプログラムモジュールであり、GCスレッド11はこれらのモジュールを呼び出すことによって、GCを実行させる。また、インタプリタ12は、この例で、「スレッド4」からオブジェクト生成依頼があったとき、GCモジュール10の「オブジェクト生成」を呼び出し、またオブジェクト間の参照関係の変更依頼があったとき、GCモジュール10の「マーク付与」を呼び出す。

【0074】なお、図4に示した例では、各スレッドが

中間コード (例えばJavaアプレットのバイナリコード) で表されているが、中間コードをVMICに対するネイティブコードに変換するコンパイラを設けてもよい。

(例えばJavaの場合、JIT [Just-In-Time コンパイラ] 等を設ける。) この場合、ネイティブコードで記述したスレッドであるので、図4に示したインタプリタ12を介さずにGCモジュール10を直接アクセスすることになる。

【0075】図5はGCを行うことによって、ヒープ領域内に生じるフラグメントを解消するためのコンパクションを行った例を示す図である。図中ハッチング部分がオブジェクトであり、これをコンパクションすることによって、(B)に示すようにフラグメントが解消されて、連続したメモリ空間が広がることになる。

【0076】上記コンパクションは図4に示したGCモジュールの「コンパクション」のプログラムにより行う。

【0077】図6はコンテキストスイッチ発生有無を検出するためのAPIの使用例を示す図である。同図の(A)に示すように、処理1を実行する前にコンテキストスイッチ発生有無の検出開始を依頼するAPI #Aを発行してから処理1を実行する。処理1の終了後、コンテキストスイッチ発生有無検出の終了を依頼するAPI #Bを発行する。(A)に示す例では、同様にコンテキストスイッチが発生していないので、そのまま次の処理2を行う。もし、(B)に示すように、処理1の途中でスレッド#2の処理が行われれば、すなわちコンテキストスイッチが発生していれば、API #Bの発行後、処理1を破棄する。たとえば領域Aの内容を領域Bにコピーするような処理で、コピー処理中にコンテキストスイッチが発生した場合、領域Aの内容が変わって領域AとBの内容が不一致となる場合がある。不一致ではコピーしたことにならないので、領域Bを無効とする。このことは、最初から処理が行われなかったのと同じであり、処理自体を破棄したことに他ならない。

【0078】図7は上記の処理をフローチャートとして示したものである。まず、API #Aを発行してから処理1を実行する (s1→s2)。この処理1が終了した後、API #Bを発行して、その戻り値を取得する (s3)。戻り値がコンテキストスイッチの発生を表す場合、処理1を破棄して、再び処理1を実行する (s4→s5→s1→s2)。戻り値がコンテキストスイッチの非発生を表す場合、処理を終了する。このように所定の期間内でのコンテキストスイッチの発生有無が分かるので、コンテキストスイッチが発生すれば、その間の処理を破棄し無効とすることによって、システムを現実にはロックしていないにも拘らず排他的制御を行うことが可能となる。

【0079】図8は上記API #AおよびAPI #Bのカーネルにおける処理手順を示すフローチャートであ

る。API #Aの発行（システムコール）があれば、コンテキストスイッチの発生有無を示すフラグをクリアする。また、API #Bが発行されると、上記フラグの状態を戻り値としてスレッドに返す。

【0080】図9はコンテキストスイッチの処理手順を示すフローチャートである。コンテキストスイッチがスケジューラによって行われると、上記フラグをセットしてからコンテキストスイッチを実行する。すなわちスイッチ前のスレッドの実行状態をコンテキストとして格納するとともに、スイッチ後のスレッドのコンテキストを  
10 読み出してCPUのレジスタ等に設定する。

【0081】図10は上記コンパクションの処理手順を示すフローチャートである。まず、ヒープ領域内の先頭のオブジェクトを指定し（s11）、そのオブジェクトをヒープ領域の先頭にコピーするための領域を確保し（s12）、コピー中に他のスレッドによってその領域に何らかのデータが書き込まれないようにしてから上記API #Aを発行する（s13）。そして、図5に示したように、ヒープ領域内のオブジェクトを先頭から順次空き領域へコピーすることによって詰めていく（s1  
20 4）。1つのオブジェクトについてのコピーが終われば、上記API #Bを発行する（s15）。このことにより、API #Aを発行してから、API #Bを発行するまでの期間にコンテキストスイッチが発生したか否かがAPI #Bの戻り値として得られる。もしコンテキストスイッチが発生していれば、今回コピーを行ったオブジェクトを既に確保している領域に再びコピーする（s16→s13→...）。もしコンテキストスイッチが発生していなければ、次のオブジェクトについて同様に処理を行う（s16→s17→s18→...）。この  
30 ようにシステムをロックすることなく、しかも他のスレッドとともに、コンパクションを同時に行うことが可能となる。

【0082】上述の例ではマーク&スイープ法によるGCにおけるコンパクションに適用したが、複写法によるGCに適用する場合、図11および図12に示す処理を行う。

【0083】図11は上記複写法によるGCのフローチャートである。まず、オブジェクトの参照関係を表すツリー構造のデータのルートノードへポインタを移動し（s21）、そのルートノードに相当する、From領域にあるオブジェクトをTo領域に複写する（s22）。（ヒープ領域はFrom領域とTo領域に分けられ、From領域内の残すべきオブジェクトのみをTo領域に複写することによってTo領域にガベージのないオブジェクトだけを再構成する。今回は現在のFrom領域をTo領域とし、To領域をFrom領域として、その操作を交互に繰り返すのが、複写法によるガベージコレクションである。）その後、ツリーを辿って、参照関係にある次のオブジェクトへポインタを移動させ（s  
50

23）、そのオブジェクトをTo領域に複写する（s24）。この処理をツリーで進めるすべてのオブジェクトについて行う。

【0084】図12は図11における複写処理の内容を示すフローチャートである。まず、複写すべきオブジェクトをTo領域内の所定位置にコピーするための領域を確保し、コピー中に他のスレッドによってその領域に何らかのデータが書き込まれないようにし、上記API #Aを発行してから（s31）、オブジェクトをFrom領域からTo領域に複写する（s32）。その後、上記API #Bを発行する（s33）。このことにより、API #Aを発行してから、API #Bを発行するまでの期間にコンテキストスイッチが発生したか否かがAPI #Bの戻り値として得られる。もしコンテキストスイッチが発生していれば、今回複写を行ったオブジェクトを既に確保している領域に再び複写する（s34→s31  
→...）。

【0085】図13および図14はコンテキストスイッチの発生有無を検出して排他性を確保するのではなく、複写しようとするオブジェクトが複写前に書き替えられたか否かを検出して排他性を確保する場合の処理手順を示すフローチャートである。

【0086】複写を行う場合、まず複写すべきオブジェクトをTo領域内の所定位置にコピーするための領域を確保し、コピー中に他のスレッドによってその領域に何らかのデータが書き込まれないようにしてから、図13に示すようにAPI #Cを発行する（s41）。このAPIは指定したメモリ領域ない対する書き込みが発生したか否かの検出を依頼するアプリケーションプログラムインタフェースである。続いてオブジェクトをFrom領域からTo領域に複写する（s42）。その後、上記API #Dを発行する（s43）。このAPIは上記API #Cが呼び出されてから、このAPI #Dが呼び出されるまでの間に指定メモリ領域に何らかのデータの書き込みが発生したか否かが戻り値として返されるアプリケーションプログラムインタフェースである。したがって、複写しようとするオブジェクトのメモリ領域を指定してAPI #Cを発行し、API #Dも戻り値を見れば、そのオブジェクトが参照されたか否かが判る。オブジェクトが参照されていれば、内容が書き変わっている可能性があるため、そのオブジェクトの複写をやり直す（s44→s41→...）。

【0087】図14は上記API #CおよびAPI #Dのカーネルにおける処理手順を示すフローチャートである。API #Cの発行（システムコール）があれば、所定にワークエリアをクリアし、API #Cのパラメータで指定された領域がライトされたときに例外が発生するようにMMU（Memory Management Unit）を設定する。また、API #Dが発行されると、上記パラメータで指定された領域がライトされたときに例外が発生しないよ

うに、MMUに対する上記設定を解除する。そして上記ワーク領域にセットされるフラグの状態を戻り値としてスレッドに返す。

【0088】図15は上記例外発生時のMMUの処理内容を示すフローチャートである。例外が発生すると、上記ワーク領域内の参照フラグをセットする。

【0089】なお、上述の例では複写法によるGCの場合について示したが、マーク&スイープ法におけるコンパクションの場合にも同様に適用できる。

【0090】【第2の実施形態】図16および図17はインクリメンタルにガベージコレクションできるGCスレッドの優先度を自動的に切り替えるようにした場合の例を示す図である。

【0091】図16の(A)に示すように、GCスレッドの優先度を、高優先度時間は高め(例えば最高優先度にし)、低優先度時間は低く(例えば最低優先度)する動作を交互に行う。

【0092】同図の(B)はGCスレッド以外の中程度の優先度のスレッドを同時に行った場合の例について示している。すなわちGCスレッドの優先度が低いとき、それより優先度の高いスレッドがReady状態となれば、コンテキストがスイッチされ、そのスレッドの実行中にGCスレッドの優先度が高くなれば、そのGCスレッドにコンテキストがスイッチされる。そのGCスレッドの処理が中断すると、上記のGCスレッド以外のスレッドの処理を行うことになる。このようにすれば、一定周期で必ずGCスレッドが実行されるため、フリー領域が慢性的に不足状態となることが防止され、常に高いパフォーマンスを維持することができる。

【0093】図17はスレッドの優先度の切替に関するカーネルの行う処理手順を示すフローチャートである。優先度の値は複数段階あり、その値の設定は対応するAPIの発行により行えるようにしている。ここでは、GCスレッドの2つの優先度を設定する際、高優先度の値を設定するAPIが発行されると、図17の(A)に示すように、その値をGCスレッドの高優先度の値として設定する。同様に、低優先度の値を設定するAPIが発行されると、(B)に示すように、その値をGCスレッドの低優先度の値として設定する。また、GCスレッドの高優先度の時間を設定するAPIが発行されると、同図の(C)に示すように、その値を設定し、同様に低優先度の時間を設定するAPIが発行されると、同図の(D)に示すように、その値を設定する。

【0094】図17の(E)は、カーネルのスケジューラに対する処理手順を示すフローチャートである。ここでは、初期状態で、高優先度のキューにGCスレッドが資源の割当てを受けようとする待ち行列に入っているものとし、まず、そのデータ(GCスレッドを識別するデータ)を高優先度のキューから取り出して低優先度のキューに挿入する(s51)。そしてスケジューラは、既に

設定されている低優先度時間だけ時間待ちを行い(s52)、続いて低優先度のキューからGCスレッドを識別するデータを取り出して高優先度のキューに挿入する(s53)。そしてスケジューラは、既に設定されている高優先度時間だけ時間待ちを行う(s54)。以上の処理を繰り返すことによって、スケジューラの動作により図16(A)に示したようにGCスレッドの優先度が交互に切り替わる。GCスレッド以外のスレッドについては、そのスレッドの優先度に対応するキューの待ち行列に応じて従来と同様のスケジューリングが行われ、図16(B)に示したように、コンテキストスイッチがなされる。

【0095】図18は上記のGCスレッドの高優先度時間をAPIによって設定可能とした場合について示している。また、図19はそのAPIの呼び出しに応じたカーネルにおける処理手順を示すフローチャートである。このAPIの発行(システムコール)があれば、上記の高優先度時間をそのAPIのパラメータによって設定(登録)する。

【0096】図20は上記のGCスレッドの高優先度時間を設定可能とするAPIを使うプログラムの例を示すフローチャートである。まず、ループの1回目の示すフラグの状態を見る(s61)。最初はOFF状態であるので、このフラグをONし(s62)、現在のフリー領域を調べ、それを記録する(s63)。初めはGCの高優先度時間と低優先度時間は予め定められたデフォルト値である。スケジューラにより、一定時間停止した後(s64)、今度は、前回に調べたフリー領域の容量と現在のフリー領域の容量とを大小比較を行い(s65)、フリー領域の容量が増加していれば、GCスレッドの高優先度時間を短くし(s66→s67)、フリー領域の容量が減少していれば、GCスレッドの高優先度時間を長くする(s68)。以上の処理を繰り返す。このことによって、GCの優先度が動的に自動調整される。

【0097】図21はGCスレッドの高優先度時間と低優先度時間とによる周期を設定可能とした場合について示している。(A)は周期が長い場合、(B)は短い場合である。

【0098】図22は上記周期を設定可能とする周期設定APIの呼び出しに応じたカーネルにおける処理手順を示すフローチャートである。この周期設定APIの発行があれば、現在設定されている高優先度時間および低優先度時間の値を読み取り、その割合(比率)を計算する(s71→s72→s73)。その後、この周期設定APIのパラメータで指定された周期の値に応じて高・低優先度時間を再計算する(s74)。そして、その高優先度時間および低優先度時間を設定更新する(s75→s76)。

【0099】たとえば連続的なパルスを処理するよう

な、連続してCPU資源が必要なアプリケーションプログラムの場合には、GCスレッドの周期が短くなるように周期設定APIを発行し、通常のアプリケーションプログラムのように断続的にCPU資源を利用するアプリケーションプログラムでは周期が長くなるように周期設定APIを発行する。

【0100】〔第3の実施形態〕図23および図24は必要な時点でGCスレッドを実行し、且つリアルタイム性を確保するようにした例であり、図23に示す例では、スレッド1とスレッド3がリアルタイム性の要求されるスレッド、スレッド2がリアルタイム性の要求されないスレッド（ノーマルスレッド）である。また、スレッド3がGCスレッドである。メモリのフリー領域が多い通常状態で、スレッド2が実行されているときにイベント1が発生すれば処理がスレッド1に移され、イベント1の処理に起因するスレッド1の処理が終了すればスレッド2へ戻る。同様に、イベント2が発生すればスレッド3に処理が移る。もしスレッド2の処理によってフリー領域が予め定めた警告レベルまで低下したとき、スレッド2の処理を中断して、スレッド3のGCスレッドを実行する。この処理によってフリー領域が確保されるとスレッド2の処理へ戻る。もしGCスレッドの処理途中からイベント1が発生すれば、GCの途中であってもスレッド1へ処理が移る。このようにリアルタイム性の要求されるスレッドでは、一般に生成されるオブジェクトの量が少なく、大体の予測が可能であるので、それに応じて上記警告レベルを設定しておけば、スレッド2に処理によって、フリー領域が大幅に減少してスレッド1やスレッド3のリアルタイム性の要求される処理が実行できなくなるといった、不都合が回避できる。

【0101】図24は上記コンテキストスイッチの処理を行うスケジューラの処理手順を示すフローチャートである。まずイベントの発生有無を検知（s81）、イベントが発生していれば、対応するリアルタイムスレッドにコンテキストをスイッチする（s82）。イベントが発生していなくて、フリー領域が警告レベルより多ければ、ノーマルスレッドのうち優先度の最も高いもの（図23に示した例ではスレッド2）にコンテキストをスイッチする（s83→s84）。もしフリー領域が警告レベルより少なくなれば、GCスレッドにコンテキストをスイッチする（s85）。

【0102】〔第4の実施形態〕図25はコンテキストスイッチの検出を依頼するAPIにおける強制コンテキストスイッチの例を示している。先に示した例では、API #Aを発行してから、API #Bを発行するまでの間にコンテキストスイッチが発生したか否かが判るようにしたものであったが、この図25に示す例は、GCの優先度を高・低交互に切り替える場合に、上記の排他制御のためのAPIを発行している間にコンテキストスイッチが発生することを予測して、無駄な処理を行わない

ようにしたものである。すなわち、高優先度GCスレッドを実行中にAPI #A'を発行した際、API #A'は高優先度時間が終了するまでに、必要な処理が完了するか否かを判定して、もし完了しなければ、その時点でGCスレッドの優先度を低くする。図25に示した例では、GCスレッドの優先度が低くなったことにより、GCスレッド以外の中優先度のスレッドにコンテキストスイッチされることになる。このことにより無駄な処理が避けられる。

【0103】図26は上記API #A'の呼び出しに応じたカーネルにおける処理手順を示すフローチャートである。このAPIの呼び出しがあれば、コンテキストスイッチフラグをクリアし（s91）、GCスレッドの高優先度時間の残時間を取得する（s92）。そして、このAPI #A'のパラメータである排他時間（API #A'を発行してから、API #Bを発行するまでの時間）と上記残時間との大小比較を行う（s93）。そして、残時間が排他時間より短ければGCスレッドの優先度を強制的に低くする（s94）。なお、API #B呼び出しによる処理内容とコンテキストスイッチの処理内容は図8および図9に示したものと同様である。

【0104】〔第5の実施形態〕図27はメモリ（ヒープ領域）使用量に応じてGCのアルゴリズムを切り替えるための処理手順を示すフローチャートである。まず、メモリ使用量の値とそれに応じてどのアルゴリズムでGCを行うかを示すしきい値を設定し（s101）、メモリ使用量を取得する（s102）。これはヒープ領域内に生成されたオブジェクトのメモリサイズの合計値である。メモリ使用量がしきい値th1を超えた場合、GCアルゴリズム（1）の手順でGCを行う（s103→s104）。メモリ使用量がしきい値th2を超えたなら、GCアルゴリズム（2）の手順でGCを行う（s105→s106）。以下同様である。ここで、しきい値th1はしきい値th2より小さく、GCアルゴリズム（1）としては図16に示した、GCスレッドの優先度の高低を交互に繰り返す処理を行う。しきい値の高いときに行うGCアルゴリズムとしては、図23に示した不定期のGCを実行する。ここでGC自体は例えばマーク&スイープ法により行う。通常、前者の場合にはCPU負荷が軽く、専らこの方法によりGCが行われるが、ノーマルスレッドによって短時間に大量のメモリが使用された場合には、図23に示した方法によりGCを重点的に行う。このことによって常に広いフリー領域を確保し、且つリアルタイム性を維持する。

【0105】〔第6の実施形態〕図28～図30はオブジェクト生成時のヒープ領域に対する新たなオブジェクトの割り当てサイズを定めるための処理に関する図である。一般にプログラムの実行により生成されるオブジェクトのサイズの分布は図28に示すように略正規分布を示す。その中心値は32と64バイトの間に納まる程度

である。そこで、図29の(A)に示すように、この中心値より大きなサイズで、且つ2の巾乗バイトの整数倍のサイズをオブジェクトの割り当てサイズとする。従来は同図の(B)に示すように、生成すべきオブジェクトのサイズの量だけ任意に割り当てられていたため、そのオブジェクトの消去の際、不揃いのサイズのフラグメントが生じる。本願発明によれば、新たに生成されるオブジェクトのサイズが上記固定値の整数倍であるため、メモリ領域の再利用性が高まり、メモリ使用効率が全体に高まる。しかも場合によってはコンパクションが不要となる。

【0106】図30はそのオブジェクト生成時の処理手順を示すフローチャートである。先ず、これまでに生成したオブジェクトのサイズの発生頻度の分布データを求める(111)。既に前回までに分布データを得ている場合には、それを更新する。続いて、今回割り当てるときメモリの空いている領域で、且つ生成しようとするオブジェクトのサイズより大きな領域を探し、上記の固定サイズの整数倍のメモリ領域にオブジェクトを割り当てる(112→113→114→115)。上記2の巾乗バイトはシステム定数であるが、必ずしもこの値を固定サイズとする必要はなく、任意である。固定サイズをもし大きく過ぎる値に採れば、サイズの大きな領域に小さなオブジェクトが割り当てられる場合が増えることになり、逆に、固定サイズを小さく過ぎる値に採れば、オブジェクトの生成に再利用できない領域が増えることになる。分布データを基に上記固定サイズを決定する場合には、メモリの使用効率が最適となるように決定すればよい。また、最適値でなくても、例えば2の巾乗の値を探れば、アドレス決定がやり易くなるという効果を奏する。

【0107】図31は上記オブジェクトのサイズの発生頻度分布データを求めるプログラムモジュール、およびオブジェクトの割り当てサイズを決定するプログラムモジュールの処理内容を示すフローチャートである。先ず、オブジェクトのサイズの発生頻度分布データを求めるプログラムモジュールをロードし(121)、そのプログラムモジュールを起動する。すなわち一定時間が経過するまで(たとえばアプリケーションプログラムの起動・終了が10回行われるまで、またはたとえば24時間経過するまで)、生成された各オブジェクトのサイズをサイズ毎に計数し、その分布データを求める(122→123)。その後、その分布データをシステムに登録し(124)、オブジェクトのサイズの発生頻度分布データを求めるプログラムモジュールをアンロードする(125)。続いて固定サイズを決定するプログラムモジュールをロードし(126)、そのプログラムモジュールを実行する。すなわちシステムの登録された分布データを取得し、その分布中心値より大きなサイズで、且つたとえば2の巾乗バイトを固定サイズとして

決定し、その値をシステムに登録する(127→128)。その後、固定サイズを決定するプログラムモジュールをアンロードする(129)。

【0108】このように、一度オブジェクトのサイズの分布が検知された後、および固定サイズが決定された後は、それらのためのプログラムモジュールをアンロードすることによって、メモリ領域とCPUパワーの無駄を無くす。

【0109】〔第7の実施形態〕図49は上記固定サイズをAPIによって設定する例を示している。図に示すように、固定サイズ設定処理では、固定サイズを引数として固定サイズ設定APIを呼び出す。呼び出されたAPIでは、引数を固定サイズとしてシステムに登録する。オブジェクト生成時にはオブジェクトのサイズと固定サイズを比較し(211)、固定サイズ以下であれば、ヒープ上の固定サイズの空き領域を探し、見つかった領域をオブジェクトに割り当てる(212→213)。固定サイズを超える場合は、ヒープ上のオブジェクトサイズより大きな空き領域を探し、見つかった領域をオブジェクトに割り当てる(214→215)。

【0110】〔第8の実施形態〕図50は上記固定サイズをAPIによって設定する他の例を示している。図50に示すように、この例では、まずオブジェクトサイズの分布データをオブジェクトサイズ分布設定APIを呼び出して設定する。この分布データは予め所定時間所定のアプリケーションを実行して測定したものである。オブジェクトサイズ分布設定APIは呼び出しに応じて、引数をオブジェクトサイズ配列変数にセットする。続いて、その中心値より大きなサイズで、且つ2の巾乗バイトを固定サイズとして決定し、これを固定サイズ変数にセットする。

【0111】〔第9の実施形態〕図51～図53はオブジェクトの割り当てサイズを予めいくつかのサイズに定めておく例を示す図である。図51は事前に所定のアプリケーションを所定時間実行して測定したオブジェクトサイズの分布および分割領域の集合を表している。すなわちヒープ領域を分割する場合、実際のオブジェクトサイズの分布を測定し、その分布に近くなるように、分割サイズと数を定める。たとえば、ヒープ領域が2MBの場合、分割サイズ指定変数を

とする。

【0112】上記分割サイズを設定する場合、図52に示すように、分割サイズ設定APIを呼び出して行う。呼び出された分割サイズ設定APIは図53に示すよう

集合No.	n	バイトk	個数m
1	64	5000	
2	256	10000	
3	1k	10000	
4	4k	5000	
5	32k	500	



に、引数を分割サイズ指定変数にセットし、それに応じてヒープ領域を分割する。すなわち、まずループカウンタ $n$ を0とし( s 2 2 1 )、 $n$ 番目の分割サイズ指定変数からサイズ $k$ と個数 $m$ を取得する( s 2 2 2 )。続いてヒープ領域からサイズ $k$ の領域を $m$ 個に分割し(  $m$ 個分確保し)、リストに登録する( s 2 2 3 )。この処理をループカウンタ $n$ をインクリメントしながら繰り返す、全てのサイズの分割を行う( s 2 2 5→s 2 2 2→...)。なお、上記の例で、サイズが32kバイトを超えるオブジェクトを生成する場合は、ヒープ領域内の上記分割された領域以外の領域に割り当てる。

【0113】【第10の実施形態】図32～図34はオブジェクトの寿命を考慮してGCの効率を高めるための処理を行う例を示す図である。図32の(A)に示す例では、ヒープ領域として、短寿命のオブジェクトを生成する領域と長寿命のオブジェクトを生成する領域とに分け、クラスは長寿命領域に確保する。尚、クラスはその他の固定領域に確保してもよい。そして、クラスにはオブジェクトを生成するためのテンプレートとしての定義情報以外に、生成するオブジェクトの寿命を示す寿命フラグを持たせる。この寿命フラグはクラスの生成時に自動的に生成する。同図の(B)は従来のヒープ領域に対するオブジェクトの割り当て例を示す図である。

【0114】図33はオブジェクトの消去の手順を示すフローチャートである。同図に示すようにオブジェクトを消去しようとする際、そのオブジェクトの寿命が長いと短いとを判定し、寿命が短ければ、そのオブジェクトのクラスの寿命フラグを短寿命にセットする。例えばオブジェクトの領域内にそのオブジェクトが生成された時刻を格納しておき、そのオブジェクトを消去する時刻との差によって、そのオブジェクトの寿命を求める。上記時刻は例えばGCの回数を単位としてもよい。

【0115】図34はオブジェクトの生成の処理手順を示すフローチャートである。クラスの寿命フラグが短寿命を示していれば、短寿命領域にオブジェクトを生成し、そうでなければ、長寿命領域にオブジェクトを生成する。

【0116】このようにオブジェクトの寿命に応じてメモリの割り当て領域を区分することによって、長寿命領域ではフラグメントを大幅に低下でき、メモリの使用効率が向上する。また、例えば寿命領域についてGCを重点的に行うことにより、GCに消費されるCPUパワーを小さくすることができる。

【0117】【第11の実施形態】次に、マーク&スイープ法によるインクリメンタルGCについて図35～48を参照して説明する。

【0118】図35はマーク&スイープ法によるGCの全体の処理手順を示すフローチャートである。このGCはマークテーブルのクリア、上記ツリー探索によるマーク付与およびオブジェクトの消去(スイープ)を繰り返

し行う。

【0119】図36は図35における「マークテーブル」の処理内容を示すフローチャートである。この処理はマークテーブルの内容を一旦クリアするものであり、まずマークテーブルの先頭にポインタを移動し( s 1 3

1)、その位置のマークをクリアし( s 1 3 2)、次のマークの位置までポインタを移動させる( s 1 3 3)。この処理を全てのマークについて繰り返す( s 1 3 4→s 1 3 2→...)。

【0120】図37は図35における「オブジェクトの消去」の処理内容を示すフローチャートである。まず、マークテーブルの先頭にポインタを移動させる( s 1 4 1)、マークの有無を検出し、マークされていないならば、そのマークテーブル上の位置に相当するオブジェクトのヒープ領域内の位置を計算し、該当のオブジェクトを消去する( s 1 4 2→s 1 4 3→s 1 4 4)。続いて、マークテーブルのポインタを次に移動して、同様の処理を繰り返す( s 1 4 5→s 1 4 6→s 1 4 2→...)。これによりマークテーブルにマークされているオブジェクトを残し、その他のオブジェクトをヒープ領域から消去する。

【0121】説明を容易にするために、先ず前提となるマーク&スイープ法におけるマーク付与について説明する。

【0122】図38はツリー探索によるマーク付与の手順を示す図である。(A)に示すように、ルートノード10から各ノードへ、ツリー構造で表される参照関係を辿って、参照関係にあるノード(オブジェクト)にマークを付与する。具体的にはマークテーブル上の該当位置のビットをセットする。このツリー構造は、たとえ他のオブジェクトが他のどのオブジェクトを参照しているかを示す、オブジェクト内に設けられる変数の内容によって構成され、このオブジェクトの参照関係を辿ることが、すなわちツリーを辿ることである。

【0123】(A)のように、ノード番号3までマーク付与を行った時点で割込が発生し、その割込処理によって、(B)のように、ルートノードからノード番号7で表されるオブジェクトへの参照関係が絶たれて、ノード番号2で表されるオブジェクトからノード番号7で表されるオブジェクトが参照される関係となれば、割込処理が終了してGCスレッドに戻って、マーク付与を再開したとき、ルートノードからノード番号7で表されるオブジェクトへの参照関係が絶たれているので、(C)に示すようにポインタをルートノードに戻して次の参照関係にあるノード番号8に進むことになる。この時点ではノード番号5、6に対してはマーク付与されない。そこで、参照関係の変更のあったオブジェクトについては、そのオブジェクトからツリーを辿ってそのオブジェクトから参照されているオブジェクトについてマークを付与する必要がある。

【0124】図39は「オブジェクトの生成」の処理内容を示すフローチャートである。まずカーネルに対してシステムロックを起動し(s151)、ヒープ領域内の空いている領域を探し(s152)、生成しようとするオブジェクトのサイズより大きな領域に対して必要なサイズを割り当て(s153→s154)、参照変更のマーク付与(ライトバリア)を行い(s155)、システムをアンロックする(s156)。

【0125】図40は上記「参照変更のマーク付与」の処理手順を示すフローチャートである。まず、参照変更されたオブジェクトからマークテーブル上の位置を計算し、該当のマークがWhiteであるかを判定する。このWhiteマークはたとえば00の2ビットで表され、未だマークされていない状態を意味する。もし、Whiteマークでなければ、すでにマークされているので、そのまま処理を終了する。Whiteマークであれば、それをGrayにマークする。このGrayマークはたとえば01の2ビットで表され、参照変更のあったオブジェクトであることを意味する。なお、オブジェクトからマークの位置を計算する際、たとえばオブジェクトのアドレスを1/8してオフセットを加えることによって行うか、オブジェクトの通し番号により計算する。

【0126】図41は「ツリー探索によるマーク付与」の処理手順を示すフローチャートである。まず、ツリーを辿るためのポインタをツリーのルートノードに移動させ(s161)、新たに生成するオブジェクトに対応するマークを付与する(s162)。続いてツリーを辿って、ポインタを次のオブジェクトへ移動させ(s163)、この処理をツリーの最後まで繰り返す(s164→s162→...)。その後、各スレッドスタックの先頭へポインタを移動させ(s165)、スタックにあるオブジェクトに対応するマークを付与する(s166)。その後、ポインタをスタックの次に移動させて(s167)、この処理をツリーの最後まで繰り返す(s168→s166→...)。その後、スレッドスタックの次に移り(s169)、同様の処理をスレッドスタックの最後まで繰り返す(s170→s166→...)。さらにこのスレッドスタックについての処理をすべてのスレッドスタックについて行う(s171→s172→s165→...)。この一連のツリー探索において、Grayマークが途中で1度でも検出された場合、もう一度ルートノードから探索およびマーク付与を行う(s173→s161→...)。

【0127】図42は図41における「オブジェクトに対応するマーク付与」の処理手順を示すフローチャートである。この処理は、生成されているオブジェクトからマークテーブル上の位置を計算し、Blackマークを付与する。このBlackマークはたとえば1xの2ビットで表され、マークされている状態を意味する。

【0128】上述したようなマーク付与の方法では、そ

の途中で割込がかかってオブジェクトの参照関係が変化すると、Grayマークが付与されるので、ツリー探索を繰り返さなければならない。場によってはいつまでもマーク付与が完了せずに、GCがいつまでも行われなといった事態に陥る。

【0129】図43は上記の問題を解消するための「ツリー探索によるマーク付与」の手順を示す図である。

(A)は図38に示した(A)の状態で割込がかかって参照関係が変化したときの状態を示す。このように、ノード番号3までマーク付与を行った時点で割込が発生し、その割込処理によって、ルートノードからノード番号7で表されるオブジェクトへの参照関係が絶たれて、ノード番号2で表されるオブジェクトからノード番号7で表されるオブジェクトが参照される関係となれば、参照先のノード番号7表すデータをマークスタックに積む。その後、割込処理が終了してGCスレッドに戻って、マーク付与を再開したとき、ルートノードからノード番号7で表されるオブジェクトへの参照関係が絶たれているので、(B)に示すようにポインタをルートノードに戻して次の参照関係にあるノード番号8をマークする。この時点で一通りのツリー探索を終了し、その後は、(C)に示すようにマークスタックの内容によって示されるノードからツリーを辿って参照関係にあるオブジェクトについてマークを付与する。これにより(D)に示すように、参照関係にある全てのオブジェクトについてのマーク付与が完了する。

【0130】図44は「参照変更のマーク付与」の処理手順を示すフローチャートである。このように、参照変更されたオブジェクトを示すデータを上記マークスタックに積む。なお、オブジェクトの生成の処理は図39に示したもの変わらない。

【0131】図45は、「ツリー探索によるマーク付与」の処理手順を示すフローチャートである。ステップs161～s172の部分は図41に示したフローチャートのステップs161～s172と同じである。この図45に示す例では、全スレッドスタックについてのツリー探索を完了した後、マークスタックからデータを取り出し(s181→s182)、そのデータで示されるオブジェクトに対応するマーク付与を行い(s183)。そのオブジェクトからツリーを辿ってツリーの最後まで、参照関係にあるオブジェクトのマーク付与を行う(s184→s185→s183→...)。この処理をマークスタックが空になるまでマークスタックのポインタを更新しながら繰り返す(s181→s182→...)。

【0132】図46は図45における「オブジェクトに対応するマーク付与」の処理手順を示すフローチャートである。この処理は、生成されているオブジェクトからマークテーブル上の位置を計算し、マークを付与する。

【0133】このようにマークスタックを用いることに

よって、ツリー探索をルートノードから再開する必要があるなくなり、マーク付与に要する全体の処理時間を大幅に短縮化できる。

【0134】図47および図48は上記マークスタックを用いてマーク付与を行う場合に、更に無駄な処理時間を無くすようするフローチャートである。図47は上記「参照変更のマーク付与」の処理手順を示すフローチャートである。まず、参照変更されたオブジェクトからマークテーブル上の位置を計算し、該当のマークがWhite 10 であるか否かを判定する(s191→s192)。このWhite マークは上述したように未だマークされていない状態を意味する。もし、White マークでなければ、すでにマークされているので、そのまま処理を終了する。White マークであれば、それをGrayにマークする(s193)。上述したようにこのGrayマークは参照変更のあったオブジェクトであることを意味する。続いてマークスタックに参照先のオブジェクトを示すデータを積み(s194)。

【0135】図48は上記「オブジェクトに対応するマーク付与」の処理手順を示すフローチャートである。この処理は、生成されているオブジェクトからマークテーブル上の位置を計算し、Black マークを付与する。このBlack マークは上述したように、マークされている状態を意味する。なお、「オブジェクトの生成」の処理は図39に示したものと同様である。

【0136】このように、参照変更のあったオブジェクトについてマークを付与する場合、そのオブジェクトが初めて検出されたオブジェクトである場合にのみマークを付与するようにしたため、マークスタックの内容によるツリー探索に要する時間およびマークスタックの読み出しに要する時間を短縮化できる。

【0137】なお、参照変更のあったオブジェクトについてのマークを記憶するのはスタックでなくともよく、FIFOのバッファを用いてもよい。

【0138】また、実施形態ではマークテーブルにマークを付与する際に、オブジェクトの内部にマーク用の情報を設けて、オブジェクトに直接マークを付与するようにしてもよい。

【0139】

【発明の効果】請求項1、18、25に係る発明によれば、コンピュータ資源をロックのメカニズムとして使用せずに、処理の排他性が保証される。すなわち指定メモリ領域に対するデータの書き込み有無検出の開始を依頼するAPIの呼び出しから指定メモリ領域に対するデータの書き込み有無検出の終了を依頼するAPIの呼び出しまでの間で行われたスレッドAの処理の途中で、他のスレッドによる指定メモリ領域に対する書き込みがあったか否かが、そのスレッドAで分かる。書き込みが行われていなければ、指定メモリ領域の排他性が保たれている。もし書き込みが行われていれば、スレッドAのその

間の処理を無効にして、例えばその処理を再度実行するなどの方法によって高い応答性を保ちながら、排他制御を行うことができるようになる。

【0140】請求項2、19、26に係る発明によれば、GCスレッドの優先度を交互に切り替えることによって、GCスレッドの優先度が低い状態では、他のスレッドによるアプリケーションが優先され、動作するアプリケーションが無い場合にはGCスレッドが自動的に動かされ、メモリのフリー領域が自動的に拡大される。GCスレッドの優先度が高い状態では、他のスレッドは実行されないが、優先度の低い状態では、他のスレッドにより、GCが継続して行われることによってフリー領域が慢性的に不足状態となることが防止され、常に高いパフォーマンスを維持することができる。

【0141】請求項3に係る発明によれば、APIを呼び出す側で、たとえばシステムの応答性を重視する場合には優先度の高い状態の時間を短くし、システム全体のスループットを重視する場合には優先度の高い状態の時間を長くする、といった設定が可能となり、システムのニーズに応じて性能を変えることができるようになる。

【0142】請求項4に係る発明によれば、たとえば連続的なパルス処理するような、連続してCPU資源が必要なアプリケーションプログラムの場合には、周期を短くし、通常のアプリケーションプログラムのような断続的にCPU資源を利用するアプリケーションプログラムでは周期を長くすることによって、コンテキストスイッチなどのオーバーヘッドを小さくすることができる。

【0143】請求項5に係る発明によれば、動的にガベージコレクションの優先度の高い状態の時間を調整することになり、どのようなアプリケーションプログラムでもフリー領域が足りなくなるといった状況を回避することができ、且つフリー領域が十分確保できるシステムやアプリケーションプログラムでは、ガベージコレクションを最低限に抑えることができる。

【0144】請求項6、20、27に係る発明によれば、リアルタイム性の要求されないスレッドがオブジェクトを生成して、メモリのフリー領域の量が例えばリアルタイム性の要求されるスレッドが生成するオブジェクトの量の近傍にまで減少した時に、その時点でスケジューリングが行われ、GCスレッドが実行されるため、フリー領域が直ちに確保され、リアルタイム性の要求されるスレッドの実行可能な環境を常に維持することができる。

【0145】請求項7、21、28に係る発明によれば、フリー領域または使用領域の量に応じてGCの手順を切り替えることによって、常に効率的なGCが可能となる。

【0146】請求項8、22、29に係る発明によれば、多くのオブジェクトが先に使用されていたメモリ領域を再利用できることになり、このことにより、コンパ

クションを行わなくてもメモリの使用効率が向上することになる。また、コンパクションのためのCPUパワーが不要となり、小規模のCPUで応答性の高いシステムを構成することができる。

【0147】請求項9, 10に係る発明によれば、一度オブジェクトのサイズの分布が検知された後は、また、オブジェクトの割り当てサイズが決定された後は、メモリ領域とCPUパワーの無駄がなくなる。

【0148】請求項11, 12, 30に係る発明によれば、多くのオブジェクトがメモリのヒープ領域を無駄なく再利用できることになり、コンパクションを行わなくてもメモリの使用効率が向上する。また、コンパクションのためのCPUパワーが不要となり、小規模のCPUで応答性の高いシステムを構成することができる。

【0149】請求項13, 14, 31, 32に係る発明によれば、別の装置、システム、およびアルゴリズムで、メモリのヒープ領域内に割り当てられるオブジェクトのサイズの分布を測定して、決定したオブジェクトの割り当てサイズの基になる固定サイズを登録することができ、CPUパワーとメモリの使用効率を向上させることができる。

【0150】請求項15, 23, 33に係る発明によれば、寿命データに基づいて、異なったヒープ領域にオブジェクトを生成することによって、長寿命のオブジェクトと短寿命のオブジェクトが異なった領域に生成されるため、長寿命領域でのフラグメントが大幅に低下し、メモリの使用効率が向上する。また、GCを行う場合に、短寿命のオブジェクトが生成される領域を重点的に処理することによって、GCに消費されるCPUパワーを低減することができるようになる。

【0151】請求項16, 24, 34に係る発明によれば、スweepやマークテーブルのクリアを行うときに、他のスレッドを止めなくても、スweep中またはマークテーブルのクリア中に他のスレッドにより新たなオブジェクトが生成されたり、オブジェクト間の参照関係が変更されても、マークテーブルを基にスweepを行う際に、上記の新たに生成されたオブジェクトを消去してしまうことがない。そのため、インクリメンタルにGCを行うことができ、また、自由に中断ができるので、リアルタイム性が向上する。また、常にバックグラウンドでGCスレッドを動作させておくことが可能となり、メモリの使用効率を常に高く維持できる。しかも、ツリー探索によるマーク付与に要する時間が短縮化され、割込によって、いつまでもマーク付与が完了しない、といった不都合が防止でき、GCを確実に実行できる。

【0152】請求項17に係る発明によれば、オブジェクトの参照関係の変更時の参照先が既にマークされている場合に、第2のデータの探索を重ねて記憶することがなくなり、第1のデータの探索およびマークに要する時間をその分短縮化できる。

# 【図面の簡単な説明】

【図1】実施形態に係る装置のハードウェアの構成を示すブロック図

【図2】同装置のソフトウェアの構成を示すブロック図

【図3】オブジェクト間の参照関係を示すツリーおよび各スレッドのスタックとの関係を示す図

【図4】ソフトウェアの機能ブロック図

【図5】コンパクションの作用を説明するための図

【図6】コンテキストスイッチの有無によるスレッドの処理内容の変化の例を示す図

【図7】排他制御の処理手順を示すフローチャート

【図8】コンテキストスイッチ発生有無検出のAPIに関する処理手順を示すフローチャート

【図9】コンテキストスイッチの処理手順を示すフローチャート

【図10】コンパクションの処理手順を示すフローチャート

【図11】複写法によるGCの処理手順を示すフローチャート

【図12】複写法GCにおける排他制御の処理手順を示すフローチャート

【図13】複写法GCにおける他の排他制御の処理手順を示すフローチャート

【図14】図13における排他制御用APIに関する処理手順を示すフローチャート

【図15】図13における排他制御用APIに関する処理手順を示すフローチャート

【図16】GCスレッドの優先度の自動切替の例を示す図

【図17】スレッドの優先度値および優先度時間の切替に関するフローチャート

【図18】GCスレッドの高優先度時間を変更した例を示す図

【図19】GCスレッドの高優先度時間を変更可能とするためのフローチャート

【図20】GCスレッドの高優先度時間を自動変更する例を示すフローチャート

【図21】GCスレッドの高・低優先度時間の切り替え周期を変更した例を示す図

【図22】GCスレッドの高・低優先度時間の切り替え周期を変更するためのフローチャート

【図23】不定期なGC処理の例を示す図

【図24】図21に対応するフローチャート

【図25】排他制御APIによるGCスレッドの優先度の強制変更の例を示す図

【図26】排他制御APIによるGCスレッドの優先度を強制変更するためのフローチャート

【図27】GCのアルゴリズムを切り替える処理手順を示すフローチャート

【図28】生成されるオブジェクトのサイズの分布の例

を示す図

【図 29】オブジェクトのメモリの割り当てサイズの例を示す図

【図 30】オブジェクト生成の処理手順を示すフローチャート

【図 31】オブジェクトサイズの分布検知および固定サイズの決定処理に関するフローチャート

【図 32】ヒープ領域の構成を示す図

【図 33】オブジェクト消去の手順を示すフローチャート

【図 34】オブジェクト生成の手順を示すフローチャート

【図 35】マーク&スweep法によるGCの手順を示すフローチャート

【図 36】マーククリアの処理手順を示すフローチャート

【図 37】オブジェクト消去の処理手順を示すフローチャート

【図 38】ツリー探索によるマーク付与の例を示す図

【図 39】図 38に対応するフローチャート

【図 40】図 38に対応するフローチャート

【図 41】図 38に対応するフローチャート

【図 42】図 38に対応するフローチャート

【図 43】ツリー探索によるマーク付与の例を示す図

【図 44】図 43に対応するフローチャート

【図 45】図 43に対応するフローチャート

【図 46】図 43に対応するフローチャート

【図 47】参照変更のマーク付与の他の処理手順を示すフローチャート

【図 48】オブジェクトに対応するマーク付与の他の処理手順を示すフローチャート

10 【図 49】固定サイズの設定とオブジェクト生成に関する処理手順を示すフローチャート

【図 50】オブジェクトサイズ分布設定に関する処理手順を示すフローチャート

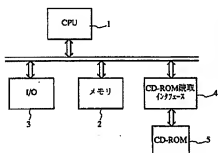
【図 51】オブジェクトサイズの分布と分割サイズの例を示す図

【図 52】ヒープ領域を所定サイズで分割する処理とオブジェクト生成に関する処理手順を示すフローチャート

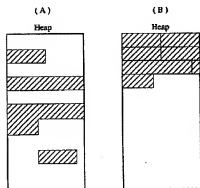
【図 53】ヒープ領域を所定サイズで分割する処理に関する処理手順を示すフローチャート

20 【図 54】従来のマーク&スweep法によるGCの手順を示すフローチャート

【図 1】



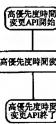
【図 5】



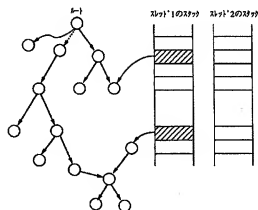
【図 2】



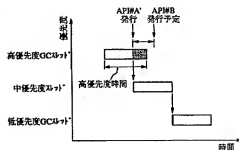
【図 19】



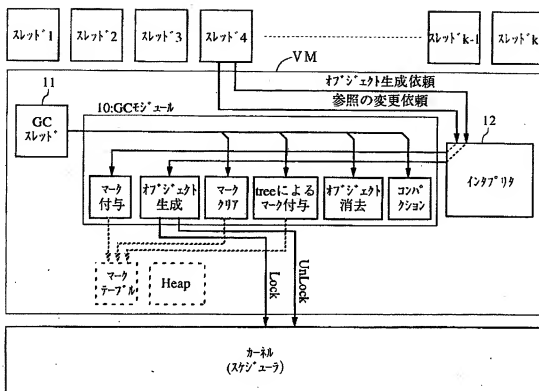
【図 3】



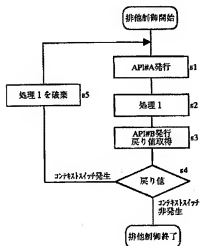
【図 25】



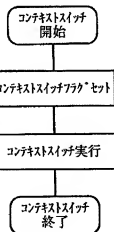
【図 4】



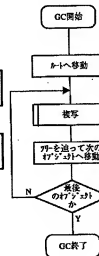
【図 7】



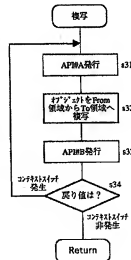
【図 9】



【図 11】

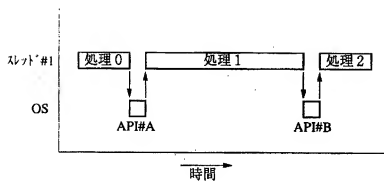


【図 12】

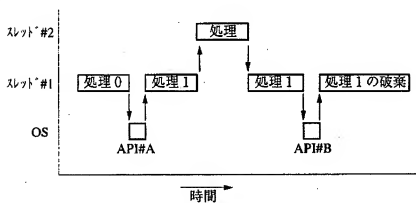


【図 6】

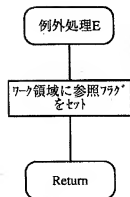
( A )



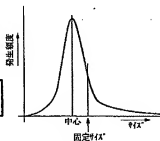
( B )



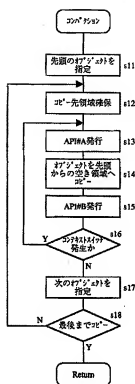
【図 15】



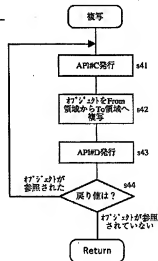
【図 28】



【図 10】



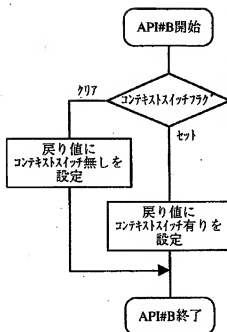
【図 13】



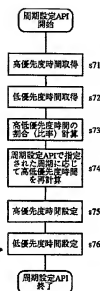
【図 8】



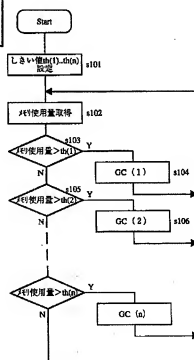
(B)



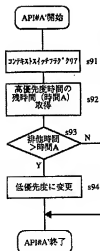
【図 22】



【図 27】



【図 26】

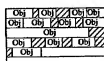


【図 29】

(A)

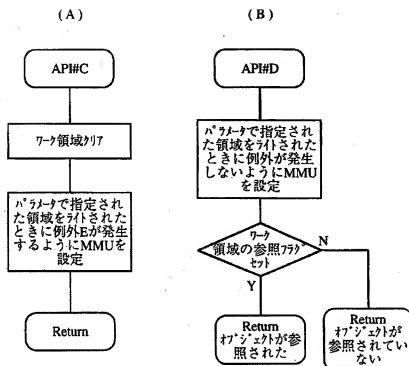


(B)

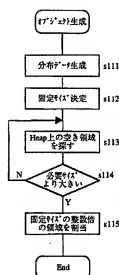




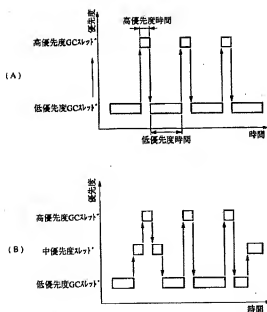
【図 14】



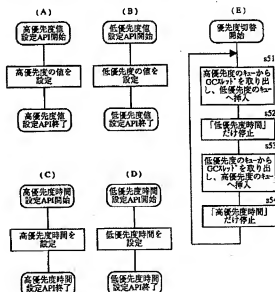
【図 30】



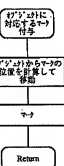
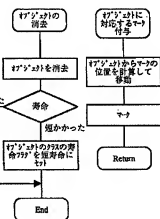
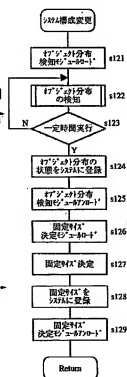
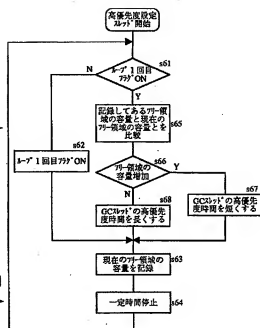
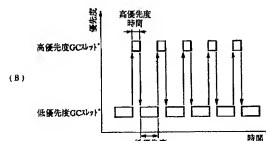
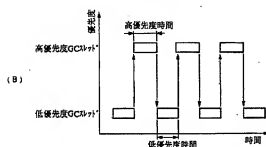
【図 16】



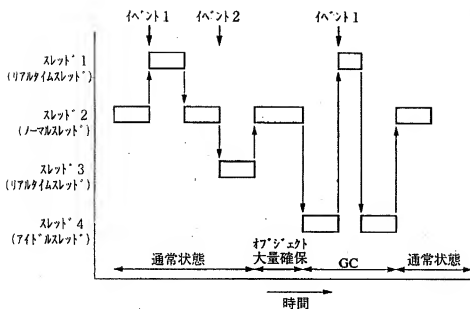
【図 17】



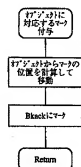
【图 20】



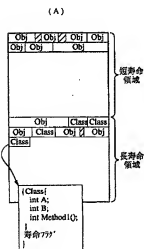
【図 23】



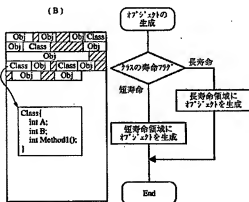
【図 48】



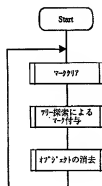
【図 32】



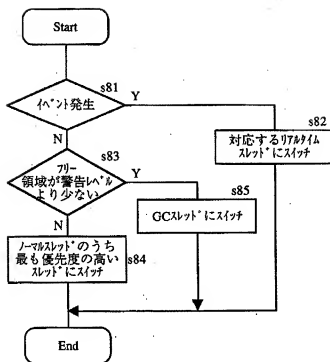
【図 34】



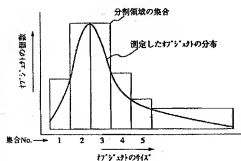
【図 35】



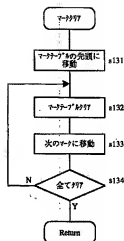
【図24】



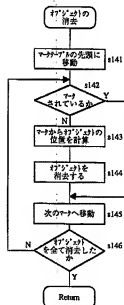
【図51】



【図36】

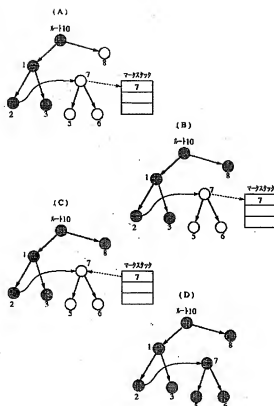


【図37】

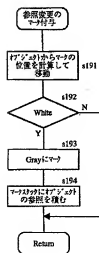




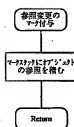
【図 4 3】



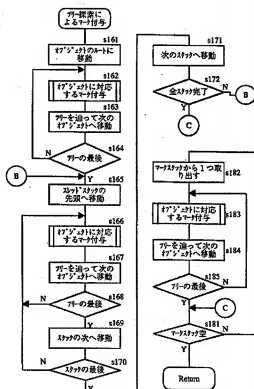
【図 4 7】



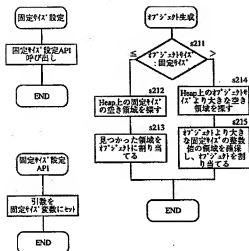
【図 4 4】



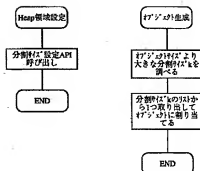
【図 4 5】



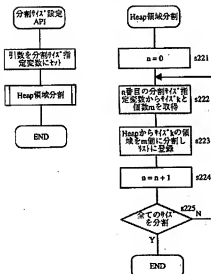
【図49】



【図52】



【図53】



フロントページの続き

(72)発明者 小中 義治

京都府京都市右京区花園土堂町10番地 オ  
ムロン株式会社内

(72)発明者 栗林 博

京都府京都市右京区花園土堂町10番地 オ  
ムロン株式会社内